Estimator & Panel Data Model.

Def. Extremum Estimators / M-estimators: $\hat{\theta}$

estimators that max or min some function of the sample data.

$$\hat{\theta} \quad \text{solves} \quad \min_{\theta \in \Theta} / \max \quad \hat{Q}_n(\theta)$$

$\hat{Q}_n(\theta)$: sample objective function.

θ_0: population parameter of interest.

e.g. MLE, GMM

Def. Population: with individual characterized by (Y, X, U) where $Y(X, U) \sim$ distribution P.
Observations. Y_i, X_i $i = 1, \cdots, n$. assume (Y_i, X_i) i.i.d $\sim P_{Y|X}$.

Consider the problem of predicting Y given X.

i) Suppose we know $P_{Y|X}$.

Let $h(x)$ be a prediction for Y given $X = x$. Suppose you incur loss of $(Y - h(x))^2$ for your prediction error. Choose $h(\cdot): Supp(X) \to \mathbb{R}$ to minimize $L(h) = \mathbb{E}[(Y - h(x))^2]$

$\to$ solved by $h(x) = \mathbb{E}[Y | X = x]$

ii) Now suppose $h(\cdot)$ is constrained to be linear, i.e. $h(x) = xb$ for some column vector b.

Then if β solves $\min_b \boxed{\mathbb{E}[(Y - Xb)^2]}$ $Q(b)$

$\to$ OLS β, the best linear predictor for Y given X.

under mild regularity conditions. $\beta = \mathbb{E}[X'X]^{-1} \mathbb{E}[X'Y] = [\int X'X \, dP_X]^{-1} [\int X'Y \, dP_X]$

With data we don't know $P_{Y|X}$ and we don't know $\mathbb{E}[\ ]$, we can estimate β by

$$\hat{\beta}_{OLS} = \left(\frac{1}{n} \sum_{i=1}^{n} x_i' x_i\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} x_i' Y_i\right)$$

$$\hat{\beta}_{OLS} \quad \text{solves} \quad \min_{b \in B} \boxed{\frac{1}{n} \sum_{i=1}^{n} (Y_i - x_i b)^2} \quad \hat{Q}_n(b)$$

Def. A model is a set of assumptions on the data generating process.

- Economic Theory: e.g. revealed preference or equilibrium assumptions.
- Functional form assumption.
- Distributional assumptions on unobserved variables.

Data entails observed realizations of variables generated by the economic model.

e.g. $Y = X\beta + \underbrace{U}_{\text{unobserved heterogeneity}}$ and $\mathbb{E}[X'U] = 0$. All are different assumptions.

or $\underbrace{\mathbb{E}[U|X]=0}$, i.e. $\forall x,\ \mathbb{E}[U|X=x]=0$

then $\mathbb{E}[Y|X=x]=x\beta$

$$\frac{\partial \mathbb{E}[Y|X=x]}{\partial x_k}=\beta_k$$

or $U\sim N(0,\sigma^2),\ U\perp\!\!\!\perp X$

Important points: Consistency: $\hat{\theta}_n \xrightarrow{P} \theta$; Unbiasedness; Efficient; Asymptotic Variance; Asymptotic Normality.

Correct Specification: model assumptions are correct.

Stochastic Convergence Concepts.

Def. Algebraic Limit.

Let T_n be a non-random sequence of numbers. T_n converges to T as $n\to\infty$ iff
$\forall \varepsilon>0,\ \exists N_\varepsilon$, s.t. $\forall n>N_\varepsilon,\ \|T_n-T\|\le\varepsilon$.
i.e. $\lim\limits_{n\to\infty} T_n=T$. $T_n\to T$ as $n\to\infty$.

Def. Convergence in Probability.

Let T_n be a sequence of random variables. T_n converges in probability to T iff
$\forall \varepsilon>0,\ \lim\limits_{n\to\infty}\mathbb{P}(\|T_n-T\|>\varepsilon)=0$
i.e. $T_n\xrightarrow{P}T$. $\operatorname*{plim}\limits_{n\to\infty} T_n=T$, or $\|T_n-T\|=o_p(1)$.

Def. Almost Sure Convergence.

T_n converges a.s. to T iff $\mathbb{P}(\lim\limits_{n\to\infty}\|T_n-T\|=0)=1$
i.e. $T_n\xrightarrow{a.s.}T$, or $\|T_n-T\|=o_{a.s.}(1)$.

Def. Convergence in L_d norm:

T_n converges in L_d norm to T iff
$\lim\limits_{n\to\infty}\mathbb{E}[|T_{nj}-T_j|^d]=0$. $\forall j=1,\cdots,\dim(T)$. i.e. $T_n\xrightarrow{L_d}T$
if $d=2$, then $T_n\xrightarrow{M.S.}T$ converges in mean square.

Theorem. $T_n\xrightarrow{M.S.}c$ iff $\lim\limits_{n\to\infty}\mathbb{E}[T_n]=c$ and $\lim\limits_{n\to\infty}\operatorname{var}(T_n)=0$.

Def. Convergence in distribution:

Let T_n have CDF F_n, for each $n=1,2,\cdots$. T_n converges in distribution to T iff
$\lim\limits_{n\to\infty}F_n(t)=F(t)$ for every t at which F is continuous.

This does not require $\|T_n-T\|$ to be small!! e.g. $Z\sim N(0,1)$, $X=-Z$, X also $\sim N(0,1)$.

Let $X_i: i=1,\cdots,n$ be a random sample of draws $X\sim F_X$.

Theorem: LLN are theorems that state $\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_X)\xrightarrow{P}0$ under some given conditions.

e.g. Chebyshev's LLN: Let $\{X_i\}_{i=1}^{n}$ be a random sample of k-dimentional unrelated observations from a distribution with a common mean and variance: $\mathbb{E}[X_i]=\mu$, $var(X_i)=\Sigma\in\mathbb{R}^{k\times k}$, $dim(X_i)=k$.

Then: $\bar{X}_n\xrightarrow{P}\mu$.

Theorem: CLT. says $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right)\xrightarrow{d}Z$, $Z\sim N(0,\Sigma)$. $\mu=\mathbb{E}[X]$ under some conditions.

e.g. Lindeberg-Lévy CLT: Let $\{X_i\}_{i=1}^{n}$ be a random vector with mean $\mathbb{E}[X_i]=\mu\in\mathbb{R}^k$, variance $var(X_i)=\Sigma$.

Then, $\sqrt{n}(\bar{X}_n-\mu)\xrightarrow{d}Z$, $Z\sim N(0,\Sigma)$.

Def. The Characteristic Function (CF) of X:

$$g(t)=\mathbb{E}[\exp(itX)]=\int_{-\infty}^{+\infty}\exp(itx)f_X(x)\,dx \quad \text{where } x \text{ has density } f_X(x).$$

This is a Fourier Transformation of $f_X(x)$ density function.

$$g^{(r)}(0)=i^r\,\mathbb{E}[X^r]$$

Fourier transformation are invertible.

If $\int_{-\infty}^{\infty}|g(t)|<\infty$, then $f_X(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-i\pi x)g(t)\,dt$.

So the distribution of X can be uniquely determined & equally characterized by the CF.

Theorem: Continuity Theorem:

Let T_n each has CDF F_n and CF $g_n(t)$.

Suppose $\exists\, g(\cdot):\mathbb{R}\to\mathbb{C}$ s.t.

① $g_n(t)\to g(t)$ $\forall t\in\mathbb{R}$; ② $g(\cdot)$ is continuous at 0

Then $g(\cdot)$ is the CF of some $T\sim F(\cdot)$, and $T_n\xrightarrow{d}T$.

Useful Theories:

1. Slutsky Theorem:

Let X_n and Y_n be random vectors.

$X_n\xrightarrow{d}X$, $Y_n\xrightarrow{P}c$. Then:

① $X_n+Y_n\xrightarrow{d}X+c$ ② $Y_nX_n\xrightarrow{d}cX$

③ $Y_n^{-1}X_n\xrightarrow{d}c^{-1}X$

2. Continuous Mapping Theorem

X: random k-vector. $g(\cdot):\mathbb{R}^k\to\mathbb{R}$

continuous on a set $\mathcal{X}$ s.t $\mathbb{P}(X\in\mathcal{X})=1$

X_n: a sequence of random k vectors.

Then: ① $X_n\xrightarrow{P}X\Rightarrow g(X_n)\xrightarrow{P}g(X)$

② a.s. ③ d.

3. Delta Method:

X_n: sequence of random k vector

Z: random k vector.

$\sqrt{n}(X_n-\mu)\xrightarrow{d}Z$

Let $F:\mathbb{R}^k\to\mathbb{R}^s$ be continuously differentiable at μ. Then

$\sqrt{n}(F(X_n)-F(\mu))\xrightarrow{d}\nabla F(\mu)Z$

proof of Lindeberg-Lévy CLT.

first we calculate the CF of X where $X\sim N(0,1)$.

$$\begin{aligned}
g(t)=\mathbb{E}[\exp(itX)]&=\int_{-\infty}^{\infty}\exp(itx)\frac{1}{\sqrt{2\pi}}\exp(-\tfrac{1}{2}x^2)\,dx\\
&=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}(x^2-2itx+(it)^2-(it)^2)\right]dx\\
&=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}(x-it)^2-\tfrac{t^2}{2}\right]dx\\
&=\exp(-\tfrac{t^2}{2})\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp(-\tfrac{1}{2}(x-it)^2)\,dx\\
&\overset{y=x-it}{=}\exp(-\tfrac{t^2}{2})\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp(-\tfrac{1}{2}y^2)\,dy\\
&=\exp(-\tfrac{t^2}{2})
\end{aligned}$$

univariate case: Let $T_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)$, $Y_i=\frac{X_i-\mu}{\sigma}$, $\mu=\mathbb{E}[X_i]$, $\sigma^2=var(X_i)$

then: $\mathbb{E}[Y_i]=0$, $var(Y_i)=1$. CF: $\gamma(t)=\mathbb{E}[\exp(itY_i)]$

the CF of T_n is $g_n(t)=\mathbb{E}[\exp(it\frac{1}{\sqrt{n}}\sum Y_i)]$

$=\mathbb{E}[\prod_{i=1}^{n}\exp(it\frac{Y_i}{\sqrt{n}})]$

$$= \prod_{i=1}^{n} \mathbb{E}[\exp(it\frac{Y_i}{\sqrt{n}})] \quad \text{since } Y_i \text{ i.i.d.}$$

$$= \prod_{i=1}^{n} \psi(\frac{t}{\sqrt{n}})$$

$$= \psi(\frac{t}{\sqrt{n}})^n$$

$$\ln g_n(t) = n \ln \psi(\frac{t}{\sqrt{n}})$$

$$= n \cdot m(t)$$

Taylor expansion around $\frac{t}{\sqrt{n}} = 0$ + intermediate value theorem, take derivative w.r.t. $\frac{t}{\sqrt{n}}$:

$$\ln g_n(t) = n[m(0) + m'(0)(\tfrac{t}{\sqrt{n}} - 0) + m''(0)\frac{(t/\sqrt{n} - 0)^2}{2!} + m'''(0)\frac{(\tilde{t}/\sqrt{n} - 0)^3}{3!}] \quad \text{where } \tilde{t} \text{ is between } 0 \text{ and } t.$$

$$m(0) = \ln \psi(0) = 0$$

$$m'(0) = \psi'(0)/\psi(0) = 0$$

$$m''(0) = \frac{\psi''(0)\psi(0) - \psi'(0)^2}{\psi(0)^2} = -1$$

$$\text{So } \ln g_n(t) = -\frac{1}{2}t^2 + \frac{m'''(0)\,\tilde{t}^3}{3!\sqrt{n}}$$

$$\longrightarrow -\frac{1}{2}t^2 \quad \forall\, t \in \mathbb{R}.$$

$$g_n(t) \longrightarrow \exp(-\tfrac{1}{2}t^2) = g(t) \quad \text{pointwise.}$$

So $T_n \xrightarrow{d} X$ where $X \sim N(0,1)$.

$\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n}\frac{X_i - \mu}{\sigma}) \xrightarrow{d} X$ where $X \sim N(0,1)$.

Use Mean Value Theorem to show.

$$\nabla F(\mu) = \begin{bmatrix} \frac{\partial F_1(\mu)}{\partial x_1} & \cdots & \frac{\partial F_1(\mu)}{\partial x_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_s(\mu)}{\partial x_1} & \cdots & \frac{\partial F_s(\mu)}{\partial x_k} \end{bmatrix}$$

if $Z \sim N(0, \Sigma)$, then $\sqrt{n}(F(\bar{X}_n) - F(\mu)) \xrightarrow{d} N(0, \nabla F(\mu)\Sigma\nabla F(\mu)')$

4. Cramer-Wold Device:

$X_n \xrightarrow{d} X \iff \lambda' X_n \xrightarrow{d} \lambda' X, \ \forall \lambda \in \mathbb{R}^k$

use it to transform multivariate normal distribution into univariate.

Multivariate case: now know $\{X_i\}_{i=1}^{n}$ i.i.d. $\mathbb{E}[X_i] = \mu$, $var(X_i) = \Sigma$.

Σ is positive definite. then, normalize

$Y_i = \Sigma^{-\frac{1}{2}} X_i$. then $\mathbb{E}[Y_i] = \Sigma^{-\frac{1}{2}}\mu \in \mathbb{R}^k$, $Var(Y_i) = \Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}} = I_k$

Let $\lambda \in \mathbb{R}^k$, then $\lambda' \bar{Y}_n \in \mathbb{R}$.

$$\lambda' \bar{Y}_n = \lambda' \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n}\lambda' Y_i$$

$$\mathbb{E}[\lambda' Y_i] = \lambda' \Sigma^{-\frac{1}{2}}\mu \in \mathbb{R}, \quad var(\lambda' Y_i) = \lambda' I_k \lambda \in \mathbb{R}$$

by univariate Lindeberg-Lévy CLT:

$$\sqrt{n}(\lambda' \bar{Y}_n - \lambda' \Sigma^{-\frac{1}{2}}\mu) \overset{d}{\sim} N(0, \lambda' I_k \lambda)$$

$$\lambda' \sqrt{n}(\bar{Y}_n - \Sigma^{-\frac{1}{2}}\mu) \overset{d}{\sim} N(0, \lambda' I_k \lambda)$$

$$\lambda' \sqrt{n}(\bar{Y}_n - \Sigma^{-\frac{1}{2}}\mu) \xrightarrow{d} Z', \ Z' \sim N(0, \lambda' I_k \lambda)$$

then, by Cramer-Wold device,

$$\sqrt{n}(\bar{Y}_n - \Sigma^{-\frac{1}{2}}\mu) \xrightarrow{d} Z, \ Z \sim N(0, I_k)$$

$$\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n}\Sigma^{-\frac{1}{2}}X_i - \Sigma^{-\frac{1}{2}}\mu) \xrightarrow{d} Z, \ Z \sim N(0, I_k)$$

$$\Sigma^{\frac{1}{2}}\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n}\Sigma^{-\frac{1}{2}}X_i - \Sigma^{-\frac{1}{2}}\mu) \xrightarrow{d} \Sigma^{\frac{1}{2}}Z, \ \Sigma^{\frac{1}{2}}Z \sim N(0, \Sigma)$$

then, $\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n}X_i - \mu) \xrightarrow{d} \Sigma^{-\frac{1}{2}}Z, \ \Sigma^{\frac{1}{2}}Z \sim N(0, \Sigma)$

$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \Sigma)$.

- MLE:

If we have a model that implies $Y|X \sim f_{Y|X}(y|x;\theta)$ with (Y, X) i.i.d, then we can estimate θ by maximizing likelihood.

The MLE solves:

$$\max_{\theta \in \Theta} \sum_{i=1}^{n} \ln f_{Y|X}(y_i|x_i;\theta)$$

This is an extremum estimator with $\hat{Q}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\ln f_{Y|X}(y_i|x_i;\theta)$.

e.g. $Y = X\beta + U$, $U \perp\!\!\!\perp X$, $U \sim N(0, \sigma^2)$

$Y \mid X \sim N(X\beta, \sigma^2)$ $\quad \theta = (\beta', \sigma')$

$f_{(Y_1 \cdots Y_n) \mid (X_1 \cdots X_n)}(y_1 \cdots y_n \mid x_1 \cdots x_n; \theta) = \prod_{i=1}^{n} f_{Y \mid X}(y_i \mid x_i; \theta)$ since i.i.d.

Under conditions to apply LLN, $\hat{Q}_n(\theta) \xrightarrow{P} Q(\theta) \equiv \mathbb{E}[\ln f_{Y \mid X}(y \mid x; \theta)]$ ☆

with population distribution.

Theorem: Suppose $Q(\theta)$ is known given knowledge of $P_{Y \mid X}$ (joint distribution of observables) and suppose $Q(\theta)$ has a unique maximizer θ_0. Then it is sufficient to establish that the population parameter (say θ_0) is point identified. ? Didn't prove?

Def. θ_0 is identifiable (identified) if $\nexists\ \theta \in \Theta$, $\theta \neq \theta_0$, s.t. $dF(z; \theta) = dF(z; \theta_0)$ w.p. 1.

or. $dF(z; \theta) = dF(z; \theta_0)$ $\quad z = (Y, X)$

$f_{Y \mid X}(\cdot \mid \cdot; \theta_0) = f^{o}_{Y \mid X}(\cdot \mid \cdot)$: actual population density of $Y \mid X$

$dF(z; \theta) = f(z; \theta)$ $\quad$ Radon-Nikodym.

$= f_{Y \mid X}(y \mid x; \theta) f_X(x)$

$F(z; \ )$ is a model of known form (a collection of structures, $\theta \in \Theta$).

$F(z; \theta)$ is a structure corresponding to the distribution of z for the given θ.

e.g. Probit

$Y = \mathbb{1}[X\beta + U \geq 0]$, $X \perp\!\!\!\perp U$, $U \sim N(0, 1)$

$\mathbb{P}(Y = 1 \mid X = x) = \mathbb{P}(X\beta + U \geq 0 \mid X = x)$

$= \mathbb{P}(-U \leq X\beta \mid X = x)$

$= \Phi(x\beta)$

$\mathbb{E}[Y] = \mathbb{P}(X\beta + U \geq 0) = \mathbb{P}(U \geq -X\beta) = \Phi(X\beta)$

$X\beta = \Phi^{-1}(\mathbb{E}[Y])$ $\quad$ If $\mathbb{E}[X'X]^{-1}$ non-singular,

$\mathbb{E}[X'X]\beta = \mathbb{E}[X'\Phi^{-1}(\mathbb{E}[Y])]$ $\quad$ $\beta = \mathbb{E}[X'X]^{-1}\mathbb{E}[X'\Phi^{-1}(\mathbb{E}[Y])]$

then β identified.

Log-likelihood is $\sum_{i=1}^{n} \ln f(y_i \mid x_i; \theta)$ $\quad$ where $f(y_i, x_i; \theta) = f(y_i \mid x_i; \theta) f_X(x_i)$

$= \sum_{i=1}^{n} y_i \ln \Phi(x_i\beta) + (1 - y_i) \ln(1 - \Phi(x_i\beta))$

β = some function of the population distribution of the observables

$\Downarrow$

sufficient to be identified

Let $\theta \neq \theta_0$, so that

$\mathbb{E}[\{x'(\theta - \theta_0)\}^2] = (\theta - \theta_0)'\mathbb{E}[x'x](\theta - \theta_0) > 0$,

implying $x'(\theta - \theta_0) \neq 0$

hence $x'\theta \neq x'\theta_0$ with positive probability

So, $f(y \mid x; \theta)$

$= \Phi(x'\theta)^y (1 - \Phi(x'\theta))^{1-y}$

$\neq f(y \mid x; \theta_0)$

because of the strict monotonicity of $\Phi(\cdot)$.

So, $f(y \mid x; \theta) f_X(x) = \begin{cases} \Phi(x\beta) f_X(x) & \text{if } y = 1 \\ (1 - \Phi(x\beta)) f_X(x) & \text{if } y = 0 \end{cases}$

β (the "θ_0" here) is identified since $\mathbb{E}[\ln f(Y \mid X; \theta)]$ is strictly concave in a NBH of θ_0 and everywhere concave

$\mathbb{E}[\ln f(Y \mid X; \theta)] = \mathbb{E}[Y \ln \Phi(X\beta) + (1 - Y) \ln(1 - \Phi(X\beta))]$

? If $\mathbb{E}[X'X]$ is non-singular, and $\mathbb{E}[\ln f(Y \mid X; \theta)]$ is concave, it is globally maximized at θ_0.

identified.

$\Downarrow$

$\mathbb{E}[|\ln f(z \mid \theta)|] < \infty$.

Theorem. if θ_0 is identified then $Q(\theta) = \mathbb{E}[\ln dF(z; \theta)]$ is uniquely maximized at θ_0 on Θ.

$= \mathbb{E}[\ln f(Y \mid X; \theta) f_X(x)]$

proof: define $H(\theta, \theta_0) = \mathbb{E}\left[\log \frac{dF(z; \theta)}{dF(z; \theta_0)}\right]$

$= \mathbb{E}\left[\log \frac{f(Y \mid X; \theta)}{f(Y \mid X; \theta_0)}\right]$

$= \mathbb{E}\left[\underbrace{\mathbb{E}\left[\log \frac{f(Y \mid X; \theta)}{f(Y \mid X; \theta_0)} \,\middle|\, X\right]}_{☆}\right]$

$$\circledast \le \log \mathbb{E}\left[\frac{f(Y|X;\theta)}{f(Y|X;\theta_0)} \,\Big|\, X\right]$$

$f(Y|X;\theta) \neq f(Y|X;\theta_0)$

since $\log$ is strictly concave, then $\le$ is $<$ unless $f(Y|X;\theta) = f(Y|X;\theta_0)$.

$(\mathbb{E}[\log 1] = \log \mathbb{E}[1] = 0)$

$$= \log \int \frac{f(Y|X;\theta)}{f(Y|X;\theta_0)} f(Y|X;\theta_0)\, dy$$

$$= \log \int f(Y|X;\theta)\, dy$$

$$= 0$$

Thus $\mathbb{E}\left[\log \frac{f(Y|X;\theta)}{f(Y|X;\theta_0)}\right] \le 0$, and $\le$ is $<$ unless $f(Y|X;\theta) = f(Y|X;\theta_0)$.

since θ_0 identified, then: $\nexists\, \theta \in \Theta,\ \theta \neq \theta_0$: $f(Y|X;\theta) = f(Y|X;\theta_0)$

So, this means that: $\forall\, \theta \in \Theta,\ \theta \neq \theta_0$: $f(Y|X;\theta) \neq f(Y|X;\theta_0)$.

then, $\mathbb{E}\left[\log \frac{f(Y|X;\theta)}{f(Y|X;\theta_0)}\right] < 0$

$$\mathbb{E}[\log f(Y|X;\theta)] < \mathbb{E}[\log f(Y|X;\theta_0)]$$

which means that θ_0 is the unique maximizer on Θ (true population parameter)

• The Basic Convergence Theorem.

Recall that: $\hat{\theta}$ maximizes $\underbrace{\hat{Q}_n(\theta)}_{\text{some function of the data.}}$ on Θ.

θ_0 maximizes $\underbrace{Q(\theta)}_{\text{some function which depends on the population distribution of observed variables.}}$ on Θ.

Convergence in P: $\hat{Q}_n(\theta) \xrightarrow{P} Q(\theta) = \mathbb{E}[\dots]$

Def: $\hat{Q}_n(\cdot)$ converges uniformly in probability to $Q(\cdot)$ on Θ; $Q(\cdot): \Theta \to \mathbb{R}$

if $\mathbb{P}\{\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q(\theta)| > \varepsilon\} \to 0$ as $n \to \infty$ or $\le \varepsilon \to 1$

$\lim_{n\to\infty} \mathbb{P}\{\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q(\theta)| > \varepsilon\} = 0 \quad \forall\, \varepsilon > 0.$

$\Leftrightarrow \sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q(\theta)| \xrightarrow{P} 0$

Theorem: $Q(\cdot)$ and $\hat{Q}_n(\cdot)$ and Θ are such that:

(i). $Q(\cdot)$ is uniquely maximized at θ_0 on Θ (sufficient condition)

(ii). Θ is compact (not critical in application)

(iii). $Q(\cdot)$ is continuous on Θ (upper semi-continuous is sufficient)

(iv). $\hat{Q}_n(\cdot)$ converges uniformly in probability to $Q(\cdot)$ (use ULLN)

then $\hat{\theta} \xrightarrow{P} \theta_0$

(Theorem also goes through if $\hat{\theta}_n$ is an approximate maximizer of $\hat{Q}_n(\cdot)$ in the sense of

$\hat{Q}_n(\hat{\theta}_n) \ge \sup_{\theta \in \Theta} \hat{Q}_n(\theta) - o_p(1)$

e.g. $T_n = \underbrace{\boxed{}}_{\text{something whose asy. dist. we can characterize}} + \underbrace{\text{junk}}_{\to\, o_p(1)}$

Once we have (i) and (ii), we have two additional conditions

Aside: $A = [a_{jk}]$. $\|A\| = \left(\sum_{j,k} a_{jk}^2\right)^{\frac{1}{2}}$ (norm)

Suppose $a(Z;\theta)$ is a $J \times K$ matrix of functions of observable vector Z and θ

Uniform LLN

Lemma 2.4. If $Z_1 \cdots Z_n$ are i.i.d, Θ compact, $a(Z;\theta)$ are continuous at each θ wp. 1 (iii)
$\exists\, d(\cdot)$ s.t. $\|a(Z;\theta)\| \le d(Z)$ (iv) for all $\theta \in \Theta$, and $\mathbb{E}[d(Z)] < \infty$
Then, $\mathbb{E}[a(Z;\theta)]$ is continuous and $\sup_{\theta\in\Theta} \|\frac{1}{n}\sum_{i=1}^{n} a(Z_i;\theta) - \mathbb{E}[a(Z;\theta)]\| \xrightarrow{P} 0$ as $n \to \infty$

• Consistency for ML

Theorem 2.5. Suppose Z_i are iid with pdf $f(Z;\theta_0)$ ⇝ ML

θ_0 = some function of the distribution.

And the followings hold

(i). $\theta \neq \theta_0 \Rightarrow f(\cdot,\theta) \neq f(\cdot,\theta_0)$

equivalent to "θ_0 is identified" in ML case? How to prove "θ_0 identified"?

$\left(\exists\, \mathcal{Z} \subseteq Supp(Z) \text{ s.t. } \forall z \in \mathcal{Z},\ f(z;\theta) \neq f(z;\theta_0) \text{ with } \mathbb{P}(z\in\mathcal{Z}) > 0 \right)$
with positive probability

Identification ⇒ uniquely maximized in ML

(ii). $\theta_0 \in \Theta$, which is compact

(iii). $\ln f(Z;\theta)$ is continuous at each θ w.p. 1

(iv)* $\mathbb{E}[\sup_{\theta\in\Theta} |\ln f(Z;\theta)|] < \infty$

$|\ln f(z;\theta)| \le \sup_{\theta\in\Theta} |\ln f(z;\theta)| = d(\cdot)$ $\xRightarrow{\text{Lemma 2.4}}$ $\hat{Q}_n(\theta)$ converges uniformly in probability to $Q(\theta)$ on Θ ☆ State this!

$\lim_{n\to\infty} \mathbb{P}\left(\sup_{\theta\in\Theta} |\frac{1}{n}\sum_{i=1}^{n} \hat{Q}_n(\theta) - Q(\theta)| > \varepsilon\right) = 0, \forall \varepsilon > 0.$

then. $\hat{\theta} \xrightarrow{P} \theta$. where $\hat{\theta}$ solves $\max_{\theta\in\Theta} \mathbb{E}[\ln f(Z;\theta)]$

$Q(\theta) = \mathbb{E}[\ln f(Z;\theta)]$ here: $a(Z;\theta) = \ln f(Z;\theta)$

$\hat{Q}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ln f(Z_i;\theta)$

• Consistency for GMM

GMM set up: A model implies $\exists\, \theta_0 \in \Theta$ s.t. $\mathbb{E}[g(Z,\theta_0)] = 0$ $\quad g(\cdot,\cdot): \mathbb{R}^K \times \mathbb{R}^L \to \mathbb{R}^J$
observed variables

e.g. (Linear IV) $Y = X\beta + u$. exogenous variables $W = (X', V')$ containing some components of X and some instrumental variables not contained in X.

Assume. $\mathbb{E}[W'u] = 0 = \mathbb{E}[\underbrace{W'}_{d^W\times 1}\underbrace{(Y - X\beta)}_{1\times 1}]$ $\quad\therefore d^W \times 1$ or $J \times 1$

Then $g(\underset{Y,X,W}{Z}, \beta) = W'(Y - X\beta)$

GMM algebraic function: $\hat{Q}_n(\theta) = \left[\frac{1}{n}\sum_{i=1}^{n} g(Z_i,\theta)\right]' \hat{\Omega} \frac{1}{n}\sum_{i=1}^{n} g(Z_i,\theta)$

A number ≥ 0, measuring the deviation from $\mathbb{E}[g(Z,\theta_0)] = 0$ when choosing different θ.

positive semidefinite $J\times J$ weighting matrix. e.g. I_J

$Q_0(\theta) = \mathbb{E}[g(Z,\theta)]' \Omega\, \mathbb{E}[g(Z,\theta)]$ Choose $\hat{\Omega}/\Omega$ to make the variance of the estimator small (more efficient)

for linear IV. 2SLS $\hat{\Omega} = (\frac{1}{n}\sum_{i=1}^{n} W_i W_i')$

If $\dim(\theta) = \dim(g(Z,\theta))$, $L = J$, then we have as many moment equations as parameters to estimate.
If, in addition, $\mathbb{E}[g(Z,\theta)]$ is linear in θ, this is a $\boxed{\text{linear system}}$ of $J = L$ variables $\mathbb{E}[g(Z;\theta)] = 0$ with $J = L$ equations to solve [can use method of moments $\frac{1}{n}\sum_i g_i(Z,\theta) = 0$)

If $J > L$, the system is said to be "over-identification" $\to$ need to $\min \hat{Q}_n(\theta)$ — still can use GMM?

It could be that there is no solution to $\mathbb{E}[g(Z,\theta)] = 0$ when $J < L$

Theorem 2.6.

Let $\hat{\theta}$ minimize the GMM objective function $\hat{Q}_n(\theta) = [\frac{1}{n}\sum_{i=1}^{n} g(Z,\theta)]' \hat{\Omega} \frac{1}{n}\sum_{i=1}^{n} g(Z,\theta)$

Suppose Z_i i.i.d. and

(i). $\hat{\Omega} \xrightarrow{p} \Omega$, a positive semidefinite $J \times J$ matrix

(ii) $\Omega\, \mathbb{E}[g(Z,\theta)] = 0$ iff $\theta = \theta_0$ ($\Leftrightarrow Q_0(\theta) = 0$ iff $\theta = \theta_0$)

Identification + (i) $\Downarrow$ uniquely minimized at $\theta = \theta_0$ in GMM

guarantee (i) of basic consistency theorem

(iii) $\theta_0 \in \Theta$, which is compact. if Ω p.d. $\Leftrightarrow \mathbb{E}[g(Z,\theta)] = 0$

$\Rightarrow$ some non-singular condition under identification.

(iv) $g(Z,\theta)$ is continuous at each θ wp 1.

(v) $\mathbb{E}[\sup_{\theta\in\Theta} \| g(Z,\theta)\|] < \infty$ $\Rightarrow \sup_{\theta\in\Theta} \| \frac{1}{n}\sum g(Z,\theta) - \mathbb{E}[g(Z,\theta)]\| \xrightarrow{p} 0$ by Lemma 2.4 Uniform L.L.N.

$\Downarrow$

$\hat{Q}_n(\theta) \xrightarrow{p} Q_0(\theta)$ uniformly. by convergence of multiples.

Then: $\hat{\theta}_n \xrightarrow{p} \theta_0$

proof: verify conditions of Theorem 2.1.

Lemma 2.3 (GMM identification)

If W is positive semi-definite and, for $g_0(\theta) = E[g(z,\theta)]$, $g_0(\theta_0) = 0$ and $Wg_0(\theta) \neq 0$ for $\theta \neq \theta_0$ then $Q_0(\theta) = -g_0(\theta)'Wg_0(\theta)$ has a unique maximum at θ_0.

Proof

Let R be such that $R'R = W$. If $\theta \neq \theta_0$, then $0 \neq Wg_0(\theta) = R'Rg_0(\theta)$ implies $Rg_0(\theta) \neq 0$ and hence $Q_0(\theta) = -[Rg_0(\theta)]'[Rg_0(\theta)] < Q_0(\theta_0) = 0$ for $\theta \neq \theta_0$. Q.E.D.

Compactness

including all its limiting values of points.

(in Euclidean space):

a closed and bounded subset

e.g. $[0,1]$ compact

$\mathbb{R}$ not compact since $+\infty, -\infty$ excluded

• Consistency without Compactness

Theorem 2.7. Let $Q_0(\cdot)$ be such that

(i) $Q_0(\cdot)$ is uniquely maximized at θ_0

(ii) θ_0 is an element of the interior of a convex set Θ and $\hat{Q}_n(\cdot)$ is concave

(iii) $\hat{Q}_n(\theta) \xrightarrow{p} Q_0(\theta)$ for all $\theta \in \Theta$ $\Rightarrow$ uniform convergence.

Then $\hat{\theta}_n$ exists w.p. 1 and $\hat{\theta}_n \xrightarrow{p} \theta_0$

• Asymptotic Normality of Extremum Estimators (Basic Result)

Theorem 3.1. Suppose $\hat{\theta}_n$ maximizes $\hat{Q}_n(\theta)$ subject to $\theta \in \Theta$ and $\hat{\theta}_n \xrightarrow{p} \theta_0$. and

(i) $\theta_0 \in$ interior (Θ)

(ii) $\hat{Q}_n(\theta)$ is twice continuously differentiable in a neighborhood of θ_0 labeled $\tilde{N}$.

(iii) $\sqrt{n}\,\frac{\partial \hat{Q}_n}{\partial \theta}(\theta_0) \xrightarrow{d} Z$ where $Z \sim N(0,\Sigma)$

(iv) $\exists\, H(\theta)$ continuous at θ_0 with $\sup_{\theta\in\tilde{N}} \| \frac{d^2\hat{Q}_n(\theta)}{d\theta\, d\theta'} - H(\theta)\| \xrightarrow{p} 0$

(v). $H = H(\theta_0)$ is non-singular.

Then, $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, H^{-1}\Sigma H^{-1})$

proof: Often, we can write $\hat{Q}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} m(Z;\theta)$. Z is observed variables, $\theta \in \Theta \subseteq \mathbb{R}^k$
$m(Z,\cdot)$ is a known function of θ.

e.g. $\frac{1}{n}\sum_{i=1}^{n} \log f(y_i|x_i;\theta)$

$$\frac{\partial \hat{Q}_n(\theta)}{\partial \theta} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial m(Z;\theta)}{\partial \theta}$$

When $\hat{Q}_n(\theta)$ does not have this representation,
we can however often show that

$$\frac{\partial \hat{Q}_n(\theta)}{\partial \theta} = \frac{1}{n}\sum_{i=1}^{n} q(Z;\theta) + o_p(n^{-\frac{1}{2}})$$

$$\Rightarrow \sqrt{n}\,\frac{\partial \hat{Q}_n(\theta)}{\partial \theta} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} q(Z;\theta) + o_p(1).$$

Since $\hat{\theta}_n$ maximizes $\hat{Q}_n(\theta)$ s.t. $\theta \in \Theta$, $\hat{\theta}_n \xrightarrow{p} \theta_0$, $\theta_0 \in$ interior (Θ),
we have $\frac{\partial \hat{Q}_n(\hat{\theta}_n)}{\partial \theta} = 0$ wp $\to 1$ as $n \to \infty$.

Mean Value Theorem: $0 = \frac{\partial \hat{Q}_n(\theta_0)}{\partial \theta} + \underbrace{\frac{\partial^2 \hat{Q}_n(\tilde{\theta}_n)}{\partial\theta\,\partial\theta'}}_{\text{(ii) guarantee its existence}}(\hat{\theta}_n - \theta_0)$ for some intermediate value $\tilde{\theta}_n$ that lies on the segment from $\hat{\theta}_n$ to θ_0

$$(\hat{\theta}_n - \theta_0) = \left(\frac{\partial^2 \hat{Q}_n(\tilde{\theta}_n)}{\partial\theta\,\partial\theta'}\right)^{-1}\left(-\frac{\partial \hat{Q}_n(\theta_0)}{\partial \theta}\right) \quad \text{by (V)}$$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\left(\frac{\partial^2 \hat{Q}_n(\tilde{\theta}_n)}{\partial\theta\,\partial\theta'}\right)^{-1}\boxed{\sqrt{n}\,\frac{\partial \hat{Q}_n(\theta_0)}{\partial \theta}} \to N(0,\Sigma) \text{ by (iii)}$$

Now, we want to show $\left\|\frac{\partial^2 \hat{Q}_n(\tilde{\theta}_n)}{\partial\theta\,\partial\theta'} - H(\theta_0)\right\| \xrightarrow{p} 0$.

$$LHS \le \left\|\frac{\partial^2 \hat{Q}_n(\tilde{\theta}_n)}{\partial\theta\,\partial\theta'} - H(\tilde{\theta}_n)\right\| + \|H(\tilde{\theta}_n) - H(\theta_0)\|$$

$$\le \sup_{\theta \in \mathcal{N}^c}\left\|\frac{\partial^2 \hat{Q}_n(\theta)}{\partial\theta\,\partial\theta'} - H(\theta)\right\| + \underbrace{\|H(\tilde{\theta}_n) - H(\theta_0)\|}_{\downarrow p \; 0}$$

since $\tilde{\theta}_n \in \mathcal{N}$, a neighborhood around θ_0 (iv)

since $\tilde{\theta}_n$ is between θ_0 and $\hat{\theta}_n$, $\hat{\theta}_n \xrightarrow{p} \theta_0$, H is continuous at θ_0.

$$= o_p(1) \xrightarrow{p} 0.$$

Thus, $\sqrt{n}(\hat{\theta}_n - \theta_0) = (H + o_p(1))^{-1}\left(-\sqrt{n}\,\frac{\partial \hat{Q}_n(\theta_0)}{\partial \theta}\right) \xrightarrow{d} (H + o_p(1))^{-1} Z$, $H = H(\theta_0)$

where $(H + o_p(1))^{-1} Z \sim N(0, H^{-1}\Sigma H^{-1})$

Suppose $\hat{Q}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} q(z_i;\theta)$
Then $\sqrt{n}(\hat{\theta}_n - \theta_0) = (H + o_p(1))^{-1}\left(-\sqrt{n}\cdot\frac{1}{n}\sum_{i=1}^{n} q'(z_i;\theta_0)\right)$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} H^{-1} q'(z_i;\theta) + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi(z_i;\theta) + o_p(1), \quad \psi\text{: influence function.}$$

- Asymptotic Distribution of ML Estimators

Estimator $\hat{\theta}_n$ maximizes $\hat{Q}_n(\theta)$ s.t. $\theta \in \Theta$

where $\hat{Q}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(\theta; z_i)$

$\ell(\theta; z) = \log f(z;\theta)$

(v) (vi)

$$\frac{\partial \hat{Q}_n(\theta)}{\partial \theta} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f(z_i;\theta)}{\partial \theta}$$

$$\frac{\partial^2 \hat{Q}_n(\theta)}{\partial\theta\partial\theta'} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log f(z_i;\theta)}{\partial\theta\partial\theta'} \longrightarrow \mathbb{E}\left[\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\right]$$

F.O.C: & (v)

$$0 = \frac{\partial \hat{Q}_n(\hat{\theta}_n)}{\partial\theta} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(z_i;\hat{\theta}_n)}{\partial\theta}$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \log f(z_i;\tilde{\theta})}{\partial\theta\partial\theta'} = \frac{0 - \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(z_i;\theta_0)}{\partial\theta}}{\hat{\theta}_n - \theta_0} \quad \text{by Mean Value Theorem}$$

Rearrange.

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 f(z_i;\tilde{\theta})}{\partial\theta\partial\theta'}\right)^{-1}}_{\downarrow p} \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log f(z_i;\theta_0)}{\partial\theta}}_{\downarrow d}$$

$$\mathbb{E}\left[\frac{\partial^2 \log f(z_i;\theta_0)}{\partial\theta\partial\theta'}\right] = H \qquad N\left(0, \operatorname{var}\left(\frac{\partial \log f(z_i;\theta_0)}{\partial\theta}\right)\right)$$

since $\mathbb{E}\left[\frac{\partial \log f(z_i;\theta_0)}{\partial\theta}\right] = 0$ F.O.C

$$\operatorname{var}\left(\frac{\partial \log f(z_i;\theta_0)}{\partial\theta}\right) = \mathbb{E}\left[\frac{\partial \log f(z_i;\theta_0)}{\partial\theta}\frac{\partial \log f(z_i;\theta_0)}{\partial\theta'}\right]$$

$= \Sigma = J$: Fisher Information Matrix.

so, $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, H^{-1}JH^{-1})$

Theorem 3.3. Let $z_1, \cdots, z_n$ be i.i.d. with pdf $f(z;\theta_0)$ and

(i) if $\theta \neq \theta_0$ then $f(z;\theta) = f(z;\theta_0)$ wp > 0 (identification)

(ii) $\theta_0 \in \Theta$ compact

(iii) $\ln f(z;\theta)$ is continuous at each $\theta \in \Theta$ wp 1

(iv) $\mathbb{E}[\sup_{\theta\in\Theta} |\ln f(z;\theta)|] < \infty$

ensure that $\hat{\theta}_n \xrightarrow{p} \theta_0$

$-H = J$

(v) $\theta_0 \in$ interior(Θ)

(vi) $f(z;\theta)$ is twice continuity differentiable in θ and $f(z;\theta) > 0$ in $\mathcal{N} \equiv$ a neighborhood of θ_0

(vii) $\int \sup_{\theta\in\mathcal{N}} \left\|\frac{\partial f(z;\theta)}{\partial\theta}\right\| dz \underset{①}{<} \infty$ and $\int \sup_{\theta\in\mathcal{N}} \left\|\frac{\partial^2 f(z;\theta)}{\partial\theta\partial\theta'}\right\| dz \underset{②}{<} \infty$

① & ②: interchange $\int$ and $\frac{d}{d\theta}$.

(viii) J exists and non-singular.

(ix) $\mathbb{E}\left[\sup_{\theta\in\mathcal{N}} \left\|\frac{\partial^2 \log f(z;\theta)}{\partial\theta\partial\theta'}\right\|\right] < \infty$.

Then: $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, J^{-1})$ where $H = -J$ so that $H^{-1}JH^{-1} = J^{-1}$. $\theta = \theta_0$ (Information Equality)

H^{-1}: the Hessian form. J^{-1}: the inner product of asymptotic variance.

Lemma 3.6

If $a(z,\theta)$ is continuously differentiable on an open set $\mathcal{N}$ of θ_0, a.s. dz, and $\int \sup_{\theta\in\mathcal{N}} \|\nabla_\theta a(z,\theta)\| dz < \infty$, then $\int a(z,\theta)dz$ is continuously differentiable and $\nabla_\theta \int a(z,\theta)dz = \int [\nabla_\theta a(z,\theta)]dz$ for $\theta \in \mathcal{N}$.

proof of Information Equality. $-H = J$

First, $\forall\, \theta \in$ interior(Θ). $\frac{\partial \log f(z;\theta)}{\partial\theta} f(z;\theta) = \frac{\partial f(z;\theta)}{\partial\theta}$ (2)

$\forall\, \theta \in$ interior(Θ): $\int \frac{\partial \log f(z;\theta)}{\partial\theta} f(z;\theta)\, dz = \int \frac{\partial f(z;\theta)}{\partial\theta} dz \overset{(vii)①}{=} \frac{\partial}{\partial\theta}\int f(z;\theta)\, dz = \frac{\partial}{\partial\theta} 1 = 0$ (1)

Differentiate (1) on both sides w.r.t. θ':

$$\frac{\partial}{\partial \theta'}\left[\int \frac{\partial \log f(z;\theta)}{\partial \theta} f(z;\theta)\, dz\right] = 0$$

$$\Big\downarrow \text{(vii) ②}$$

$$\int \left[\frac{\partial^2 \log f(z;\theta)}{\partial \theta \partial \theta'} f(z;\theta) + \frac{\partial \log f(z;\theta)}{\partial \theta} \frac{\partial f(z;\theta)}{\partial \theta'}\right] dz = 0$$

$$\Big\downarrow (2)$$

$$\int \left[\frac{\partial^2 \log f(z;\theta)}{\partial \theta \partial \theta'} f(z;\theta) + \frac{\partial \log f(z;\theta)}{\partial \theta} \frac{\partial \log f(z;\theta)}{\partial \theta'} f(z;\theta)\right] dz = 0$$

$$-\int \frac{\partial^2 \log f(z;\theta)}{\partial \theta \partial \theta'} f(z;\theta)\, dz = \int \frac{\partial \log f(z;\theta)}{\partial \theta} \frac{\partial \log f(z;\theta)}{\partial \theta'} f(z;\theta)\, dz$$

If $\theta = \theta_0$. $\quad \underset{(ix)}{-} \mathbb{E}\left[\frac{\partial^2 \log f(z;\theta_0)}{\partial \theta \partial \theta'}\right] = \mathbb{E}\left[\frac{\partial \log f(z;\theta_0)}{\partial \theta} \frac{\partial \log f(z;\theta_0)}{\partial \theta'}\right]$

$$-H = J.$$

If $\theta \neq \theta_0$, $f(z;\theta) \neq f(z;\theta_0)$, can't replace $\int \cdots dz$ with $\mathbb{E}$.

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, H^{-1} J H^{-1}) = N(0, J^{-1}) \qquad J = \mathbb{E}\left[\frac{\partial \log f}{\partial \theta} \frac{\partial \log f}{\partial \theta'}\right]$$

Under correct specification. ☆

- Estimating the Variance.

  Define score: $S(z;\theta) = \dfrac{d \log f(z;\theta)}{d\theta}$, $\quad J = \mathbb{E}[S(z;\theta)\, S(z;\theta)']$

  Asymptotic Consistent Estimators for J:

  $$\hat{J}_1 = \frac{1}{n}\sum_{i=1}^{n} S(z_i;\hat{\theta}) S(z_i;\hat{\theta})' \xrightarrow{p} \mathbb{E}[S(z;\theta)\, S(z;\theta)']$$

  $$\hat{J}_2 = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log f}{\partial \theta \partial \theta'}(z;\hat{\theta}) \xrightarrow{p} -H = J$$

  $$\hat{J}_3 = \frac{1}{n}\sum_{i=1}^{n} J(X_i;\hat{\theta}) \text{ where } J(X_i,\hat{\theta}) = \mathbb{E}[S(z;\hat{\theta}) S(z;\hat{\theta})' | X_i]$$

  $\longrightarrow J$ by LIE and LLN.

- Model is misspecified if $\nexists\, \theta \in \Theta$ s.t. density of Z is $f(z;\theta)$

  Then $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, H^{-1} J H^{-1})$ can still be true. θ_0 is called the pseudo-true value.

• Asymptotic Normality of GMM

$\dim Z = p \quad \dim \theta_0 = k \quad \dim g = J$

Set-up. For some $g(\cdot,\cdot): \mathbb{R}^p \times \mathbb{R}^k \to \mathbb{R}^J$, $\mathbb{E}[g(Z;\theta_0)] = 0$ for some unknown

GMM estimator $\hat{\theta}_n$ minimizes $\hat{Q}_n(\theta) = \left[\frac{1}{n}\sum_{i=1}^{n} g(z_i;\theta)\right]' \hat{W} \left[\frac{1}{n}\sum_{i=1}^{n} g(z_i;\theta)\right] = \hat{g}_n(\theta)' \hat{W} \hat{g}_n(\theta)$

Theorem 3.4 (Asymptotic Normality for GMM)

Suppose the hypothesis of Theorem 2.6 hold with $\hat{W} \xrightarrow{p} W$ and.

(i) $\theta_0 \in \text{interior}(\Theta)$

(ii) $g(z;\theta)$ is continuously differentiable in $\bar{\mathcal{N}}$, some neighborhood of θ_0

(iii) $\mathbb{E}[g(Z;\theta_0)] = 0$, $\mathbb{E}[\|g(Z;\theta_0)\|^2] < \infty$.

(iv) $\mathbb{E}\left[\sup_{\theta \in \tilde{\mathcal{N}}} \left\|\frac{\partial g(Z;\theta)}{\partial \theta}\right\|\right] < \infty$. ($\Rightarrow$ uniform LLN).

(v). $G'WG$ is non-singular where $G \equiv \mathbb{E}\left[\frac{\partial g(z,\theta_0)}{\partial \theta}\right]$ $\quad J \times k$.

Then: $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0,\ (G'WG)^{-1}G'W\Omega \underline{W}G(G'WG)^{-1})$ $W' = W$?

where $\Omega = \mathbb{E}[g(z;\theta_0)g(z;\theta_0)'] = Var[g(z;\theta_0)]$

$\hat{g}_n(\theta) = \frac{1}{n}\sum_{i=1}^n g(z_i;\theta)$

$\hat{G}_n(\theta) = \frac{1}{n}\sum_{i=1}^n \frac{\partial g(z_i;\theta)}{\partial\theta}$

by F.O.C.: $0 = \overbrace{\hat{G}_n(\hat{\theta}_n)'\hat{W}\hat{g}_n(\hat{\theta}_n) + \hat{g}_n(\hat{\theta}_n)'\hat{W}\hat{G}_n(\hat{\theta}_n)}^{\text{same}}$ minimization + (i).

$0 = \hat{G}_n(\hat{\theta}_n)'\hat{W}\hat{g}_n(\hat{\theta}_n)$

$= \hat{G}_n(\hat{\theta}_n)'\hat{W}[\hat{g}_n(\theta_0) + \hat{G}_n(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0)]$ for some $\tilde{\theta}_n$ between θ_0 and $\hat{\theta}_n$. (ii)

$= \hat{G}_n(\hat{\theta}_n)'\hat{W}\hat{g}_n(\theta_0) + \hat{G}_n(\hat{\theta}_n)'\hat{W}\hat{G}_n(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0)$

so $\sqrt{n}(\hat{\theta}_n - \theta_0) = -\underbrace{[\hat{G}_n(\hat{\theta}_n)'\hat{W}\hat{G}_n(\tilde{\theta}_n)]^{-1}}_{\downarrow p\ (G'WG)^{-1}} \underbrace{\hat{G}_n(\hat{\theta}_n)'\hat{W}}_{\downarrow p\ G'W} \underbrace{\sqrt{n}\hat{g}_n(\theta_0)}_{= \frac{1}{\sqrt{n}}\sum_{i=1}^n g(z_i;\theta_0) \xrightarrow{d} N(0,\Omega)}$

$\xrightarrow{d} N(0, (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1})$

pick $\hat{W}$: $\hat{W} = I_J \xrightarrow{p} W = I_J$

- Asymptotic Variance.

Theorem 4.5. If the hypothesis of Theorem 3.4 hold, and $g(z;\theta)$ is continuously differentiable at θ_0, and $\mathbb{E}[\sup_{\theta\in\tilde{N}}\|g(z;\theta)\|^2] < \infty$.

Then, $\hat{V} = (\hat{G}'\hat{W}\hat{G})^{-1}\hat{G}'\hat{W}\hat{\Omega}\hat{W}'\hat{G}(\hat{G}'\hat{W}\hat{G})^{-1}$

$\Omega = \mathbb{E}[g(z;\theta_0)g(z;\theta_0)']$ $\hat{\Omega} = \frac{1}{n}\sum_{i=1}^n g(z_i;\hat{\theta}_n)g(z_i;\hat{\theta}_n)'$

In order to get a smaller asymptotic variance,
$\hat{W}$ should give more weight to those that have less variance, less weight to those that have more variance.

$V = (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}$

If $W = \Omega^{-1}$, then $V = (G'\Omega^{-1}G)^{-1}G'\Omega^{-1}\Omega\Omega^{-1}G(G'\Omega^{-1}G)^{-1}$

$= (G'\Omega^{-1}G)^{-1}$.

It can be shown that $\overset{①}{(G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}} - \overset{②}{(G'WG)^{-1}}$ is positive semi-definite, which means that ① ≥ ② in matrix sense.

- 2-Step efficient GMM.

1. select $\hat{W}$ s.t. $\hat{W} \to W$ for some p.d. $J\times J$ matrix.
   Let $\tilde{\theta}_n$ be the GMM estimator obtained using weight matrix $\hat{W}$.
   Set $\hat{\Omega} = \frac{1}{n}\sum_{i=1}^n g(z_i;\tilde{\theta}_n)g(z_i;\tilde{\theta}_n)'$
2. compute $\hat{\theta}_n$ the GMM estimator obtained by using $\hat{\Omega}^{-1}$ as the weighting matrix.

This is the 2-step asymptotic efficient estimator (under some standard conditions).

Seemingly Unrelated Regression (SUR)

$$\star \begin{cases} Y_{i1} = X_{i1}\beta_1 + U_{i1} \\ Y_{i2} = X_{i2}\beta_2 + U_{i2} \\ \vdots \\ Y_{iG} = X_{iG}\beta_G + U_{iG} \end{cases}$$

For each observation unit i, (i.e. individuals, households, firms, etc.), we have G scalar outcomes.
If for each $g = 1, \cdots, G$, $\mathbb{E}[X_{ig}' U_{ig}] = 0$, then we have G linear regression models.
need $\mathbb{E}[X_{ig}' X_{ig}]$ to be non-singular for each β_g to be identified for any g.

$X_{ig} = (x_{ig1}, x_{ig2}, \cdots, x_{igk_g})$, $\beta_g = (\beta_{g1}, \beta_{g2}, \cdots, \beta_{gk_g})'$

OLS estimator for β_g, $\forall g = 1, 2, \cdots, G$ is $\hat{\beta}_{g,OLS} = \left(\sum_{i=1}^{n} X_{ig}' X_{ig}\right)^{-1}\left(\sum_{i=1}^{n} X_{ig}' Y_{ig}\right)$

Define

$$X_i = \begin{array}{c} \begin{matrix} 1\times k_1 & 1\times k_2 & & 1\times k_G \end{matrix} \\ \begin{bmatrix} X_{i1} & 0 & \cdots & 0 \\ 0 & X_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & X_{iG} \end{bmatrix} \end{array}$$

$G \times K$ matrix $K = \sum_{g=1}^{G} k_g$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_G \end{bmatrix}$$ $K \times 1$ vector.

This doesn't specify the relationship btw U_{ik} & U_{ig}.
If there is another model that specifies this relationship (such as GMM), then it should be more efficient.
SUR didn't use the relationship btw. each other equations.

Then, $\star \Rightarrow \underset{G\times 1}{Y_i} = \underset{G\times K}{X_i}\underset{K\times 1}{\beta} + \underset{G\times 1}{U_i}$ $\rightarrow$ system of G equations.

$$\underset{K\times 1}{\hat{\beta}_{SOLS}} = \left(\sum_{i=1}^{n} X_i' X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i' Y_i\right) = \begin{bmatrix} \hat{\beta}_{1,OLS} \\ \hat{\beta}_{2,OLS} \\ \vdots \\ \hat{\beta}_{G,OLS} \end{bmatrix}$$

If $\mathbb{E}[X_i' X_i]$ is non-singular, $\mathbb{E}[X_i' U_i] = 0$, then

$$\sqrt{n}(\hat{\beta}_{SOLS} - \beta) = \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} X_i' X_i\right)^{-1}}_{\downarrow p \ \mathbb{E}[X_i' X_i]} \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i' U_i}_{\downarrow d \ N(0,\, \mathbb{E}[X_i' U_i U_i' X_i])}$$

$$\xrightarrow{d} N(0,\, \mathbb{E}[X_i' X_i]^{-1}\, \mathbb{E}[X_i' U_i U_i' X_i]\, \mathbb{E}[X_i' X_i]^{-1})$$

for $G = 1$.
$$= \mathbb{E}[\mathbb{E}[X_i' U_i^2 X_i \mid X_i]]$$
$$= \mathbb{E}[X_i' \mathbb{E}[U_i^2 \mid X_i] X_i]$$ $\mathbb{E}[U_i^2 \mid X_i] = var(U \mid X) = \sigma^2$
$$= \sigma^2 \mathbb{E}[X_i' X_i]$$

then $\sqrt{n}(\hat{\beta}_{SOLS} - \beta) \xrightarrow{d} N(0, \sigma^2 \mathbb{E}[X_i' X_i]^{-1})$

An additional assumption sometimes made is System Homogeneity

$\mathbb{E}[U_i U_i' \mid X_i] = \mathbb{E}[U_i U_i'] = \underset{G\times G}{\Omega}$, correlation structure is not affected by X_i covariates.

$U_i = (U_{i1}, U_{i2}, \cdots, U_{iG})'$

Under System Homoskedasticity:

$\mathbb{E}[X_i' X_i]^{-1}\mathbb{E}[X_i' U_i U_i' X_i]\,\mathbb{E}[X_i' X_i] = \mathbb{E}[X_i' X_i]^{-1}\,\mathbb{E}[X_i' \Omega X_i]\,\mathbb{E}[X_i' X_i]^{-1}$ by LIE.

Under System Homogeneity and an Additional Condition, we can come up with a more efficient "GLS" estimator.

Additional Condition: $\mathbb{E}[X_i \Omega u_i] = 0$, $\mathbb{E}[X_i' \Omega^{-1} X_i]$ non-singular.

- Sufficient condition for this are:

  SGLS 1: $\mathbb{E}[X_i \otimes u_i] = 0$, i.e. $\mathbb{E}[X_{ig} u_{ih}] = 0 \quad \forall (g,h) \in \{1, \cdots, G\}^2$ strict exogeneity assumption.

  SGLS 2: $\mathbb{E}[X_i' \Omega^{-1} X_i]$ non-singular and $\mathbb{E}[u_i' u_i] = \Omega$ non-singular.

Summary of sufficient conditions for GLS.

$$\begin{cases} \mathbb{E}[X_{ig} u_{ih}] = 0 \\ \mathbb{E}[X_i' X_i] \text{ exists and non-singular.} \\ \text{Random Sampling across } i \\ \mathbb{E}[u_i u_i' \mid X_i] = \Omega \text{ is positive definite.} \end{cases}$$

then: $\hat{\beta}_{GLS} = (\sum_i X_i' \Omega^{-1} X_i)^{-1} \sum_i X_i' \Omega^{-1} Y_i$

This is the same as: define $\Omega^{-\frac{1}{2}}$ s.t. $\Omega^{-\frac{1}{2}} \Omega^{-\frac{1}{2}} = \Omega^{-1}$

$\tilde{Y}_i = \Omega^{-\frac{1}{2}} Y_i \qquad \tilde{X}_i = \Omega^{-\frac{1}{2}} X_i$

Run $\tilde{Y}_i$ on $\tilde{X}_i$ get $\hat{\beta}_{GLS}$ $\quad \tilde{u}_i = \tilde{Y}_i - \tilde{X}_i \beta = \Omega^{-\frac{1}{2}} u_i$

$$\hat{\beta}_{GLS} = (\sum_i \tilde{X}_i' \tilde{X}_i)^{-1} (\sum_i \tilde{X}_i' \tilde{Y}_i)$$

$$\sqrt{n}(\hat{\beta}_{GLS} - \beta) \xrightarrow{d} N(0, \underbrace{\mathbb{E}[\tilde{X}_i' \tilde{X}_i]^{-1} \mathbb{E}[\tilde{X}_i' \tilde{u}_i \tilde{u}_i' \tilde{X}_i] \mathbb{E}[\tilde{X}_i' \tilde{X}_i]^{-1}})$$

$$= \mathbb{E}[X_i' \Omega^{-1} X_i]^{-1} \mathbb{E}[X_i' \Omega^{-1} u_i u_i' \Omega^{-1} X_i] \mathbb{E}[X_i' \Omega^{-1} X_i]^{-1}$$

$$= \mathbb{E}[X_i' \Omega^{-1} X_i]^{-1} \mathbb{E}[X_i' \Omega^{-1} \underbrace{\mathbb{E}[u_i u_i' \mid X_i]}_{\Omega} \Omega^{-1} X_i] \mathbb{E}[X_i' \Omega^{-1} X_i]^{-1}$$

$$= \mathbb{E}[X_i' \Omega^{-1} X_i]^{-1}$$

$$\therefore \sqrt{n}(\hat{\beta}_{GLS} - \beta) \xrightarrow{d} N(0, \underbrace{\mathbb{E}[X_i' \Omega^{-1} X_i]^{-1}}_{\text{smaller than avar}(\hat{\beta}_{SOLS})})$$

- Feasible GLS, need to estimate Ω

① $\hat{\beta}_{SOLS}$, $\hat{u}_i = Y_i - X_i \hat{\beta}_{SOLS}$, use it to estimate Ω

$\hat{\Omega} = \frac{1}{n} \sum_i (Y_i - X_i \hat{\beta}_{SOLS})(Y_i - X_i \hat{\beta}_{SOLS})'$

② Then: $\hat{\beta}_{FGLS} = (\sum_i X_i' \hat{\Omega}^{-1} X_i)^{-1} (\sum_i X_i' \hat{\Omega}^{-1} Y_i)$

Reason to use GLS $\begin{cases} \text{① efficiency} \\ \text{② RE.} \end{cases}$

• Linear Panel Data Model.

$Y_{it} = X_{it}\beta + V_{it} \qquad V_{it} = c_i + u_{it}$

i: individual unit of observation. e.g. household, film, product

t: usually time, could be market, group, etc.

e.g. $\log W_t = \beta_1 + \underbrace{(\log W_{t-1})}_{\text{dynamic panel data model}}\beta_2 + eX_{it}\beta_3 + eX_{it}^2\beta_4 + ed_{it}\beta_5 + \underbrace{c_i + u_{it}}_{V_{it}}$

e.g. $\log Y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + u_{it}$ Cobb-Douglas Production Function

c_i: individual-specific time-invariant unobserved heterogeneity

u_{it}: unit and time-specific unobserved heterogeneity

In matrix form: $Y_i = X_i\beta + V_i$ where $V_i = \begin{bmatrix} c_i \\ \vdots \\ c_i \end{bmatrix} + u_i$, $u_i = \begin{bmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{bmatrix}$

$Y_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{iT} \end{bmatrix} \qquad X_i = \begin{bmatrix} X_{i1,1} & X_{i1,2} & \cdots & X_{i1,k} \\ X_{i2,1} & X_{i2,2} & \cdots & X_{i2,k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{iT,1} & X_{iT,2} & \cdots & X_{iT,k} \end{bmatrix}$. $T \times k$ matrix.

Observe for each $i=1,\cdots,n$; Y_{it}, X_{it} for $t=1,2,\cdots,T$ periods.
(Think of asymptotic as cross-sectional when T is small. When T large, over T also).

Two types of static linear panel data models: $\begin{cases} \text{Fixed Effect} \\ \text{Random Effect} \end{cases}$ distinction is assumption on the joint distribution of c_i and X_i

• Fixed Effects.

Assumption $\underline{FE1}$: $\mathbb{E}[u_{it} \mid X_i, c_i] = 0 \ \forall t$ This is a type of "strict exogeneity" assumption.

$\Rightarrow \mathbb{E}[X_{it}' u_{is} \mid c_i] = 0 \quad \forall s, t.$

weak exogeneity only restrict dependency of u_{it} with past X_{it} e.g. $\mathbb{E}[u_{it} \mid X_{it}, X_{it-1}, \cdots, X_{i1}] = 0$.

Suppose we use OLS to try to estimate β. It will be inconsistent.

Suppose we use period t and run a linear regression of Y_{it} on X_{it}

then, $\hat{\beta}_{OLS} = \left(\sum_{i=1}^{n} X_{it}' X_{it}\right)^{-1} \sum_{i=1}^{n} X_{it}' Y_{it} = \left(\sum_{i=1}^{n} X_{it}' X_{it}\right)^{-1} \sum_{i=1}^{n} X_{it}' X_{it} \beta + \left(\sum_{i=1}^{n} X_{it}' X_{it}\right)^{-1} \sum_{i=1}^{n} X_{it}' V_{it}$

$\mathbb{E}[X_{it}' V_{it}] = \mathbb{E}[X_{it}' u_{it}] + \mathbb{E}[X_{it}' c_i] \neq 0.$

"pooled OLS"

$$\hat{\beta} = \left(\sum_{i=1}^{n} X_i' X_i\right)^{-1} \left(\sum_{i=1}^{n} X_i' Y_i\right) = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} X_{it}' X_{it}\right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} X_{it}' Y_{it}\right)$$ stack all time period to run regression.

Idea. Manipulate $y_{it} = x_{it}\beta + c_i + u_{it}$ into equations that don't contain c_i

Approach ①. "Within Transformation"

$\bar{Y}_i = \bar{X}_i \beta + c_i + \bar{u}_i$ where $\forall z_{i1} \cdots z_{iT}$, $\bar{z}_i = \frac{1}{T} \sum_{t=1}^{T} z_{it}$.

then, $Y_{it} - \bar{Y}_i = (X_{it} - \bar{X}_i)\beta + u_{it} - \bar{u}_i$

$\ddot{Y}_{it} = \ddot{X}_{it} \beta + \ddot{u}_{it}$

$FE.1 \Rightarrow \mathbb{E}[\ddot{u}_{it} \mid \ddot{X}_{it}] = 0 \Rightarrow \mathbb{E}[\ddot{X}_{it}' \ddot{u}_{it}] = 0 \quad \forall t.$

$\hat{\beta}_{FE}$ is the pooled OLS estimator obtained by pooled OLS to regress $\ddot{Y}_{it}$ on $\ddot{X}_{it}$.

$$\hat{\beta}_{FE} \equiv \hat{\beta}_{WG} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \ddot{X}_{it}' \ddot{X}_{it}\right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \ddot{X}_{it}' \underbrace{\ddot{Y}_{it}}_{\ddot{X}_{it}\beta + \ddot{u}_{it}}\right)$$ Consistent.

Approach ②: "First Difference"

We can write $Y_{it} - Y_{it-1} = (X_{it} - X_{it-1})\beta + u_{it} - u_{it-1}$

$\Delta Y_{it} = \Delta X_{it} \beta + \Delta u_{it}$

$\hat{\beta}_{FD}$ is obtained by regressing ΔY_{it} on ΔX_{it} using pooled OLS (consistent since c_i gone)

$$\hat{\beta}_{FD} = \left(\sum_{i=1}^{n} \sum_{t=2}^{T} \Delta X_{it}' \Delta X_{it}\right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=2}^{T} \Delta X_{it}' \Delta Y_{it}\right)$$

Note: under FE1. $\mathbb{E}[\Delta u_{it} \mid \Delta X_{it}] = 0$ so $\hat{\beta}_{FD} \xrightarrow{p} \beta$

we need $\operatorname{rank}\left(\sum_{t=2}^{T} \mathbb{E}[\Delta X_{it}' \Delta X_{it}]\right) = k$ full rank.

so that $\frac{1}{n} \sum_{i=1}^{n} \sum_{t=2}^{T} \Delta X_{it}' \Delta X_{it} \xrightarrow{p} \sum_{t=2}^{T} \mathbb{E}[\Delta X_{it}' \Delta X_{it}]$ invertible (asymptotic over n)

• Random Effects

RE1. ⓐ $\mathbb{E}[u_{it} \mid X_{it}, c_i] = 0 \quad \forall t$ (same as FE1)

ⓑ $\mathbb{E}[c_i \mid X_i] = 0$ not imposed in FE1

RE2. $\operatorname{rank}(\mathbb{E}[X_i'\Omega^{-1}X_i]) = k$ where $\Omega = \mathbb{E}[V_iV_i']_{T\times T}$

RE3. ⓐ $\mathbb{E}[u_iu_i' \mid X_i, c_i] = \sigma_u^2 I_T = \begin{bmatrix} \sigma_u^2 & 0 & & 0 \\ 0 & \sigma_u^2 & & \\ & & \ddots & \\ 0 & & & \sigma_u^2 \end{bmatrix}$ implies homoskedasticity in u_{it} and no serial correlation in u_{it}

ⓑ $\mathbb{E}[c_i^2 \mid X_i] = \sigma_c^2$

with RE1ⓑ: $\operatorname{var}(c_i \mid X_i) = \sigma_c^2$

→ Under those RE assumption Ω has a special structure

$$\Omega = \mathbb{E}[V_iV_i'] = \mathbb{E}[V_iV_i' \mid X_i] = \begin{bmatrix} \mathbb{E}[V_{i1}^2] & \mathbb{E}[V_{i1}V_{i2}] & \cdots & \mathbb{E}[V_{i1}V_{iT}] \\ \mathbb{E}[V_{i2}V_{i1}] & \mathbb{E}[V_{i2}^2] & \cdots & \mathbb{E}[V_{i2}V_{iT}] \\ \vdots & & & \\ \mathbb{E}[V_{iT}V_{i1}] & \mathbb{E}[V_{iT}V_{i2}] & \cdots & \mathbb{E}[V_{iT}^2] \end{bmatrix} = \begin{bmatrix} \sigma_c^2+\sigma_u^2 & \sigma_c^2 & \cdots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2+\sigma_u^2 & & \\ \vdots & & \ddots & \\ \sigma_c^2 & & \cdots & \sigma_c^2+\sigma_u^2 \end{bmatrix}$$

$$\mathbb{E}[V_{it}^2 \mid X_i] = \mathbb{E}[c_i^2 \mid X_i] + 2\underbrace{\mathbb{E}[c_iu_{it} \mid X_i]}_{=0} + \mathbb{E}[u_{it}^2 \mid X_i]$$

$$\mathbb{E}[c_iu_{it} \mid X_i] = \mathbb{E}[\mathbb{E}[c_iu_{it} \mid X_i, c_i]] = \mathbb{E}[c_i\mathbb{E}[u_{it} \mid X_i, c_i]] = 0$$

For $s \neq t$:

$$\mathbb{E}[V_{is}V_{it} \mid X_i] = \mathbb{E}[c_i^2 \mid X_i] + \underbrace{\mathbb{E}[c_iu_{it} \mid X_i]}_{0} + \underbrace{\mathbb{E}[c_iu_{is} \mid X_i]}_{0} + \underbrace{\mathbb{E}[u_{is}u_{it} \mid X_i]}_{0} = \sigma_c^2$$

- If we don't use the assumption of the form of Ω:

  Idea behind GLS: Transform variables by premultiplying $\Omega^{-\frac{1}{2}}$. Run pooled OLS of $\Omega^{-\frac{1}{2}}Y_i$ on $\Omega^{-\frac{1}{2}}X_i$

  i.e. $\hat{\beta} = (\sum_{i=1}^n X_i'\Omega^{-1}X_i)^{-1}\sum_{i=1}^n X_i'\Omega^{-1}Y_i$

  Feasible GLS: First estimate Ω using pooled OLS $\Rightarrow \hat{\beta}_{OLS}$, $\hat{\Omega} = \frac{1}{n}\sum_{i=1}^n (Y_i - X_i\hat{\beta}_{OLS})(Y_i - X_i\hat{\beta}_{OLS})'$

  Then: GLS with $\hat{\Omega}$. $\hat{\beta}_{FGLS} = (\sum_{i=1}^n X_i'\hat{\Omega}X_i)^{-1}\sum_{i=1}^n X_i'\hat{\Omega}^{-1}Y_i$

  Steps of FGLS estimation of the RE model:

  1) Estimate $\hat{\beta}_{POLS}$ by pooled OLS of Y_i on X_i. Obtain residuals $\hat{V}_i = Y_i - X_i\hat{\beta}_{POLS}$.
  2) Estimate $\hat{\Omega} = \frac{1}{n}\sum_{i=1}^n \hat{V}_i\hat{V}_i'$
  3) Set $\hat{\beta}_{FGLS} = (\sum_{i=1}^n X_i'\hat{\Omega}^{-1}X_i)^{-1}(\sum_{i=1}^n X_i'\hat{\Omega}^{-1}Y_i)$.

- RE estimator $\hat{\beta}_{RE}$:

  1) Estimate $\hat{\beta}_{POLS}$ by pooled OLS of Y_i on X_i. Obtain residuals $\hat{V}_i = Y_i - X_i\hat{\beta}_{POLS}$.
  2) Estimate $\hat{\Omega}$ based on the assumed structure.

  $$\sigma_c^2 = \frac{1}{T(T-1)}\sum_{t=1}^T\sum_{\substack{s\neq t\\ s=1}}^T \mathbb{E}[V_{is}V_{it}] \qquad \hat{\sigma}_c^2 = \frac{1}{nT(T-1)-k}\sum_{i=1}^n\sum_{t=1}^T\sum_{\substack{s\neq t\\ s=1}}^T \hat{V}_{is}\hat{V}_{it}' \quad \text{(Unbiased \& consistent)}$$

  $$\sigma_c^2 + \sigma_u^2 = \sigma_v^2 = \frac{1}{T}\sum_{t=1}^T \mathbb{E}[V_{it}^2] \qquad \hat{\sigma}_v^2 = \frac{1}{nT-k}\sum_{i=1}^n\sum_{t=1}^T \hat{V}_{it}^2$$

  plug them into Ω to form $\hat{\Omega}$. Under the RE assumptions $\hat{\Omega} \xrightarrow{p} \Omega$.

  3). Set $\hat{\beta}_{RE} = (\sum_{i=1}^n X_i'\hat{\Omega}^{-1}X_i)^{-1}(\sum_{i=1}^n X_i'\hat{\Omega}^{-1}Y_i)$.

  $$\hat{\beta} = (\sum_{i=1}^n X_i'\Omega^{-1}X_i)^{-1}\sum_{i=1}^n X_i'\Omega Y_i \qquad \hat{\beta} \to \beta + \underbrace{(\frac{1}{n}\sum_{i=1}^n X_i'\Omega^{-1}X_i)^{-1}(\frac{1}{n}\sum_{i=1}^n X_i'\Omega^{-1}V_i)}_{\xrightarrow{p} 0 \text{ if } \mathbb{E}[X_i'\Omega^{-1}V_i] = 0}$$

• Specification Tests.

For maximum likelihood,

Information Matrix Identity:

If the specification is right,

$$J = \mathbb{E}\left[\frac{\partial \log f(y|x;\theta)}{\partial\theta}\frac{\partial \log f(y|x;\theta)}{\partial\theta'}\right] = -\mathbb{E}\left[\frac{\partial^2 \log f(y|x;\theta)}{\partial\theta\partial\theta'}\right] = -H$$

White Specification ML specification test or Information Test.

Quadratic form in component $J_{lk} + H_{lk}$ for l, k pairs.

For GMM. $\hat{\theta}_{GMM}$ an efficient GMM estimator solves $\min_{\theta\in\Theta} n\hat{g}_n(\theta)'\hat{\Omega}^{-1}\hat{g}_n(\theta)$ $\quad \hat{\Omega} \xrightarrow{p} \mathbb{E}[g(z;\theta)g(z;\theta)']$

at $\hat{\theta}$, the objective function is $\sqrt{n}\hat{g}(\hat{\theta})'\hat{\Omega}^{-\frac{1}{2}}\hat{\Omega}^{-\frac{1}{2}}\hat{g}(\hat{\theta})\sqrt{n}$

$\sqrt{n}\hat{g}(\hat{\theta})'\hat{\Omega}^{-\frac{1}{2}} \to N(0, I)$

If $\mathbb{E}[g(z;\theta)] = 0$ ☆ is true, then $\sqrt{n}\hat{g}_n(\hat{\theta})'\hat{\Omega}^{-1}\hat{g}_n(\hat{\theta}) \xrightarrow{d} \chi^2_{J-K}$

If this doesn't hold, then reject ☆.