SUR with same regressors

$$y = X\beta + \varepsilon$$

X now is block diagonal matrix

x_i: regressor in each equation

$$X = I_M \otimes X_0$$

on the diagonal of X

M equations
N observation
each equation has N observation
panel of data which is balanced

$\hat\beta_{GLS} = \hat\beta_{OLS}$ in this case

$$E[\varepsilon\varepsilon'|X] = \underset{NM \times NM}{\Omega} = \underset{M \times M}{\Sigma} \otimes I_N$$

Ω unknown by assumption in this case

$$\hat\beta_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

did for the system in this case not for a single equation

$$= \left[(I_M \otimes X_0')(\Sigma^{-1} \otimes I_N)(I_M \otimes X_0)\right]^{-1}(I_M \otimes X_0')(\Sigma^{-1} \otimes I_N)y$$

$$(A \otimes B)(C \otimes D) = AC \otimes BD$$

remember this very convenient rule of Kronecker products

$$= \left(\Sigma^{-1} \otimes X_0'X_0\right)^{-1}\left(\Sigma^{-1} \otimes X_0'\right)y$$

$$= \left(\Sigma \otimes (X_0'X_0)^{-1}\right)\left(\Sigma^{-1} \otimes X_0'\right)y$$

$$= \left[I_M \otimes (X_0'X_0)^{-1}X_0'\right]y$$

$$= \begin{bmatrix} (X_0'X_0)^{-1}X_0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (X_0'X_0)^{-1}X_0 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} (X_0'X_0)^{-1}X_0'y_1 \\ \vdots \\ (X_0'X_0)^{-1}X_0'y_M \end{bmatrix} = \begin{bmatrix} \hat\beta_1 \\ \hat\beta_2 \\ \vdots \\ \hat\beta_M \end{bmatrix}$$

GLS estimator in this particular case reduce to the OLS estimator

$$\mathrm{Var}(\hat{\beta}_{OLS} \mid X) = (X'\Omega^{-1}X)^{-1}$$

$\hat{\beta}_{OLS}$ $KM \times 1$ vector; $(KM \times KM)$ big matrix

in the above calculation it has already been reduced to

$$\mathrm{Var}(\hat{\beta}_{OLS} \mid X) = [\Sigma \otimes (X'X_0)^{-1}]$$

absolutely similar to the variance of the OLS with one equation $\hat{\sigma}^2_\varepsilon (X'X)^{-1}$

to compute the variance we need to estimate Σ

$$\hat{\Sigma} = \frac{1}{N} E'E$$

E: horizontally concatenate the residuals of each equation

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

$$\mathrm{Var}(\hat{\beta} \mid X) = \left[\begin{array}{c|c} \hat{\sigma}_1^2 (X_0'X_0)^{-1} & \hat{\sigma}_{12} (X_0'X_0)^{-1} \\ \hline \hat{\sigma}_{12} (X_0'X_0)^{-1} & \hat{\sigma}_2^2 (X_0'X_0)^{-1} \end{array}\right]$$

MAXIMUM LIKELIHOOD ESTIMATION

look Newton Raphson method for fun

$\{y_i\}_{i=1}^{N}$ sample of observation of random variable y iid

from distribution $f(y_i; \theta)$

Example sample of $y_i \overset{iid}{\sim} \exp(\lambda)$

sample extracted from same distribution, exponential of parameter λ

maximize the likelihood that the sample is generated by a distribution with parameter λ

I do not know the parameters but I should know the family of the distribution from which the samples are created

we can also have

important assumption

$$y_i \mid x_i \overset{iid}{\sim} N(x_i\beta; \sigma^2)$$

Gaussian linear regression

$$y_i = x_i'\beta + \varepsilon_i \qquad \varepsilon_i \sim N(0; \sigma^2_\varepsilon)$$

conditional density, otherwise we call it only density

$$f(y_1, y_2, y_3, \dots, y_N) = \prod_{i=1}^{N} f(y_i; \theta)$$

joint density

density depends on a set of unknown coefficients

$f(y_i, \lambda) = \lambda \exp(-\lambda y_i)$ this is the PDF

$\Pi[\lambda \exp(-\lambda y_i)]$

$$f(y_1 \dots y_N) = \prod_{i=1}^{N} [\lambda \exp(-\lambda y_i)]$$

density is a positive object

$$\ln[f(y_1 \dots y_N)] = \sum_{i=1}^{N} [\ln(\lambda \exp(-\lambda y_i))]$$

logarithm of the density for one single observation

log likelihood for the generic i-th observation

$$= \sum_{i=1}^{N} \log \lambda - \sum_{i=1}^{N} \lambda y_i = N \log(\lambda) - \sum_{i=1}^{N} \lambda y_i = \sum_{i=1}^{N} l_i(y_i; \lambda)$$

$$L(y_i, \lambda) = \sum_{i=1}^{N} l_i(y_i, \lambda)$$

Dividing this by N i obtain the expected log likelihood

$$E_N[l_i(y_i; \lambda)] = \sum_{i=1}^{N} \frac{l_i(y_i, \lambda)}{N}$$

Average of log likelihood for each individual observation

Maximum likelihood principle

$$\hat{\lambda}_{ML} = \underset{\lambda \in L}{\operatorname{argmax}}\, E_N[l_i(y_i; \lambda)]$$

subset of $\mathbb{R}^+$ in this case

$\underset{n \times 1}{\theta}$ then $\hat{\theta}_{ML} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, E_N[l_i(y_i, \theta)]$

1) $\hat{\theta}_{ML} \xrightarrow{P} \theta_0$? under which condition it is consistent

2) $\sqrt{N}(\hat{\theta}_{ML} - \theta_0) \xrightarrow{d} N(0, \Sigma_\theta)$?

what are we actually measuring with the average log likelihood ?

Assumptions

1) the model is correctly specified

$\{y_i\}_{i=1}^{N} \sim F(y_i;\theta)$ the distribution that I have is actually the correct distribution

$\exists$ θ_0 true parameter value such that the sample y_i that are observed are generated by that specific distribution with this parameter

Expected loglikelihood function

$$E_{\theta_0}\left[\ell_i(y_i;\theta)\right] = \int_{S(y)} \ell_i(y_i;\theta) f(y_i;\theta_0)\,dy$$

$S(y) \longrightarrow$ support of y

↳ expectation is taken with respect to the true parameter value

1) $E_{\theta_0}\left[\sup_{\theta\in\Theta}\left|\ell_i(y_i;\theta)\right|\right] < \infty$

the supremum of the expected loglikelihood function is finite

Hence the maximum has to be bounded

2) Identification

$\theta_0 \neq \theta_1$ two set of parameters are different

$\Downarrow$

$$E_{\theta_0}\left[\ell_i(y_i;\theta_0)\right] \neq E_{\theta_0}\left[\ell_i(y_i;\theta_1)\right]$$

the expected loglikelihood cannot be the same

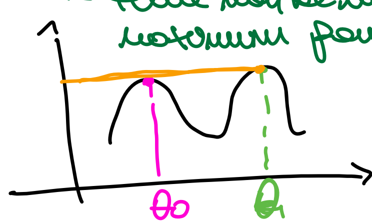
or $f(y_i;\theta_0) \neq f(y_i;\theta_1)$

density in θ_0 has to be different from density in any other point

↓

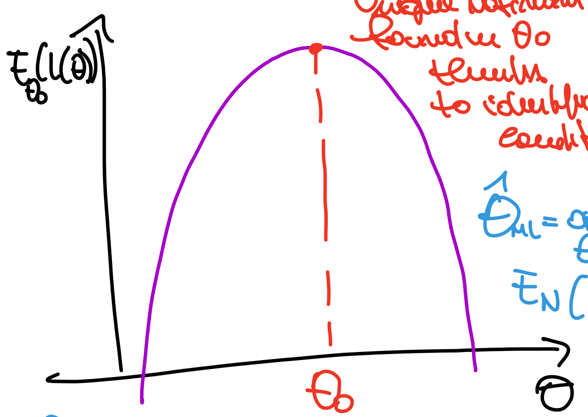
Maximum loglikelihood must be unique, otherwise there would be a problem

↳ there may be multiple maximum points



$$E_{\theta_0}\left[\ell_i(y_i;\theta_0)\right] > E_{\theta_0}\left[\ell_i(y_i|\theta)\right] \quad \forall\theta$$

Important property



Unique maximum found in θ_0 thanks to identification condition

$\hat{\theta}_{ML} = \underset{\theta\in\Theta}{\arg\max}\ E_N(\ell(\theta))$

↳ empirical analogue

$\theta_0 = \underset{\theta}{\arg\max}\ E_{\theta_0}\left[\ell(\theta)\right]$ ↳ is the purple function

Uniform law of large numbers

$$E_N(\ell(\theta)) \longrightarrow E_{\theta_0}[\ell(\theta)]$$

suppose i have y and parameter θ, h continuous function of θ

$h(y;\theta)$ 1) $E_{\theta_0}[\sup h(y;\theta)] < \infty$ → this and the identification are crucial things to check

2) $E_{\theta_0}[\sup_{\theta \in \Theta} h(y;\theta)]$ is unique

3) Θ is compact = closed and bounded

$\Downarrow$

$$E_N[h(y,\theta)] \xrightarrow{P} E_{\theta_0}(h(y,\theta)) \quad \forall \theta \in \Theta$$

As a consequence the max $\max_{\theta \in \Theta} E_N(\ell_i(y_i,\theta)) \xrightarrow{P} \max_{\theta \in \Theta} E_{\theta_0}[\ell_i(y_i;\theta)]$

↳ if it happens for any theta it has to happen for the maximum also

$\Downarrow$

$$\hat{\theta}_{ML} \xrightarrow{P} \theta_0$$

- $E_{\theta_0}\left[\frac{\partial \ell_i(y_i,\theta)}{\partial \theta}\Big|_{\theta=\theta_0}\right] = 0$ because it is the max

→ i evaluate the score in θ_0

(kx1) gradient of the function

derivative log-likelihood with respect to parameter also known as score function

↓

1° derivative loglikelihood function

INFORMATION IDENTITY

- $E_{\theta_0}\left[\frac{\partial^2 \ell_i(y_i,\theta)}{\partial\theta\partial\theta}\Big|_{\theta=\theta_0}\right]$

(k x k)

Has to be ND to have a maximum

to prove remember var as expected value outer prod | $= Var\left(\frac{\partial \ell_i(y_i;\theta)}{\partial\theta}\right)\Big|_{\theta=\theta_0}$ variance of the score

EXPECTED VALUE OF THE SCORE FUNCTION

$$\int_{-\infty}^{\infty} \frac{\partial \ln f(y_i,\theta)}{\partial \theta} f(y_i,\theta)\,dy$$

$$= \int_{-\infty}^{\infty} \frac{f'(y_i,\theta)}{f(y_i,\theta)} f(y_i,\theta_0)\,dy$$

Evaluated in θ_0

$$= \int_{-\infty}^{\infty} \frac{f'(y_i,\theta_0)}{\cancel{f(y_i,\theta_0)}} \cancel{f(y_i,\theta_0)}\,dy = \int_{-\infty}^{\infty} f'(y_i,\theta_0)\,dy$$

$$= \frac{\partial}{\partial \pi} \int_{-\infty}^{\infty} f(y_i,\theta_0)\,dy$$

$$= \frac{\partial}{\partial \pi} 1 = 0$$

because $f(y_i,\theta_0)$ is a density function

$$\int_{-\infty}^{\infty} f(y|\theta_0)\,dy = 1$$

$$\frac{\partial}{\partial\theta_0}\int_{-\infty}^{\infty} f(y,\theta_0) = 0$$

$$\Downarrow$$

$$\int_{-\infty}^{\infty} \frac{\partial f(y,\theta_0)}{\partial\theta_0}\,dy = \int_{-\infty}^{\infty} \frac{\partial \ln(f(y,\theta_0))}{\partial\theta_0}\cdot f(y|\theta_0)\,dy$$

cosa significa calcolare $E_{\theta_0}[\,\text{''}\,]_{\theta=\theta_0}$

$$E\left[\frac{\partial \ln(y_i,\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\right]$$

$$E\left[\frac{\partial\, l_i(y_i,\theta)}{\partial\theta}\right] = \int_{-\infty}^{\infty} \frac{\partial \ln(f(y_i,\theta))}{\partial\theta}\cdot f(y_i,\theta)\,dy$$

$$= \int_{-\infty}^{+\infty} \frac{\frac{\partial f(y_i,\theta)}{\partial\theta}}{f(y_i,\theta)}\cdot f(y_i,\theta_0)\,dy$$

non è evoluto in θ_0

$$= \int_{-\infty} \frac{\frac{\partial f(y_i,\theta_0)}{\partial\theta}}{\cancel{f(y_i,\theta_0)}}\,\cancel{f(y_i,\theta_0)}\,dy = \int_{-\infty}^{\infty} \frac{\partial f(y_i,\theta_0)}{\partial\theta}\,dy$$

$$= \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f(y_i,\theta)\,dy = \frac{\partial}{\partial\theta}\,1 = 0$$

+ assuming we have exponential distribution

$$f(x_i,\lambda) = \lambda e^{-\lambda x_i}$$

$$\ell(x_i,\lambda) = \ln\lambda - \lambda x_i$$

$$E_N[\ell(x_i,\lambda)] = \frac{1}{N}\sum(\ln\lambda - \lambda x_i) = \frac{1}{N} N\ln\lambda - \frac{\lambda}{N}\sum x_i$$

$$= \ln\lambda - \lambda\bar{x}$$

$$E_N\left[\frac{\partial \ell(x_i,\lambda)}{\partial\lambda}\right] = \frac{1}{\lambda} - \bar{x}$$

$$E_{\theta_0}[\ell(x_i,\lambda)] = \ln\lambda - \frac{1}{\lambda_0}$$

$$E_{\theta_0}\left[\frac{\partial \ell(x_i,\lambda)}{\partial\lambda}\right] = \frac{1}{\lambda} - \frac{1}{\lambda_0}$$

$$E_{\theta_0}\left[\frac{\partial \ell(x_i,\lambda)}{\partial\lambda}\Big|_{\lambda=\lambda_0}\right] = 0$$

max $E_N[\ln(x_i,\lambda)]$

$$= E_N\left[\frac{\partial \ell(x_i,\lambda)}{\partial\theta}\right] = \frac{1}{\lambda} - \bar{x} = 0$$

$$\hat{\lambda} = \frac{1}{\bar{x}}$$

$$E_{\theta_0}\left[\frac{\partial \ell(x_i,\lambda)}{\partial\lambda}\Big|_{\lambda=\lambda_0}\right] = \frac{1}{\lambda_0} - \frac{1}{\lambda_0} = 0$$