ECON 61001: *Lecture 10*

Alastair R. Hall

The University of Manchester

During the course have considered inference in:

- Linear regression model

- Binary response models

Developed large sample inference procedures based on assumptions about how the data are generated.

Linear regression model

- Spherical errors (& $E[x_t u_t] = 0$)
  - OLS - efficient (CS/TS)

- Non-spherical errors (& $E[x_t u_t] = 0$)
  - OLS - inefficient
    - CS: heteroscedasticity robust inference ("White's se's")
    - TS: serial correlation robust inference ("Newey-West se's")
  - GLS
    - Efficient but need model for Σ - CS/TS

Linear regression model

- $E[x_t u_t] \neq 0 \longrightarrow$ IV
  - CS: heteroscedasticity robust inference
  - TS: serial correlation robust inference

Binary response model

- LPM - OLS with heteroscedasticity robust inference
- Logit/Probit - MLE

Consistency:

$$\hat{\beta} = \beta_0 + M_T h_T$$

- M_T is a random matrix: WLLN + Slutsky $\Rightarrow M_T \xrightarrow{p} M$, finite constant
- h_T is a random vector: WLLN $\Rightarrow h_T \xrightarrow{p} 0$

Using Slutsky's Theorem,

$$\hat{\beta} \xrightarrow{p} \beta_0$$

$$T^{1/2}(\hat{\beta} - \beta_0) = M_T n_T$$

- M_T is a random matrix: WLLN + Slutsky $\Rightarrow M_T \xrightarrow{p} M$, finite constant

- $n_T = T^{1/2} h_T$ is a random vector: CLT $\Rightarrow n_T \xrightarrow{d} N(0, \Omega)$, where Ω pd, constant

$$\Rightarrow \; T^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V_\beta) \qquad \text{where } V_\beta = M \Omega M'$$

For inference need consistent estimator of V_β: use $\hat{V}_\beta = M_T \hat{\Omega} M_T'$.

Form of $\hat{\Omega}$ depends on assumptions about data.

- use knowledege of Σ i.e. conventional OLS se's or GLS
- unknown form of heteroscedasticity $\rightarrow$ White-type estimator
- unknown for of serial correlation $\rightarrow$ HAC estimator

Logit/probit models estimated via MLE.

Score equations (FOC) cannot be solved to obtain explicit formula for $\hat{\beta}$ as function of data. So proof strategy for consistency is different to OLS/GLS/IV.

But given $\hat{\beta} \xrightarrow{p} \beta_0$ can show that via first order Taylor series argument applied to score equations to show

$$T^{1/2}(\hat{\beta} - \beta_0) = M_T n_T + \xi_T$$

where

- large sample behaviour is determined by $M_T n_T$
- M_T is random matrix: WLLN + Slutsky $\Rightarrow M_T \xrightarrow{p} M$, constant
- n_T is random vector: CLT $\Rightarrow n_T \xrightarrow{d} N(0, \Omega)$, where Ω pd, constant

So similar arguments to OLS/GLS/IV

$$\Rightarrow \; T^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V_{ML}),$$

where

$$V_{ML} = M\Omega M' = \text{f(Information matrix)}$$

This generic structure,

$$T^{1/2}(\hat{\beta} - \beta_0) \;=\; M_T n_T + \xi_T,$$

holds in many nonlinear models.

- OLS based on $E[x_t u_t(\beta_0)] = 0$

- IV based on $E[z_t u_t(\beta_0)] = 0$

- MLE solves score equations

$$\left.\frac{\partial LLF_T(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}_T} = 0$$

- if data are iid then

$$\frac{\partial LLF_T(\theta)}{\partial\theta} = \sum_{t=1}^{T} \frac{\partial ln[p(v_t,\theta)]}{\partial\theta}$$

So MLE is MoM based on

$$E\left[\left.\frac{\partial ln[p(v_t,\theta)]}{\partial\theta}\right|_{\theta=\theta_0}\right] = 0 \qquad \text{(see Lecture Notes Ch 6.4)}$$

Generalized Method of Moments

So OLS, IV and MLE can all be interpreted as estimation based on the information in population moment condition.

These are all examples of a more general approach to estimation called Generalized Method of Moments (GMM).

GMM provides method to translate information about θ_0, a $p \times 1$ vector of parameters, in Population Moment Condition

$$E[f(v_t, \theta_0)] = 0,$$

into estimator of θ_0

Hansen (1982) defines the GMM estimator as:

$$\hat{\theta}_{GMM} = argmin_{\theta \in \Theta} Q_T(\theta)$$

where

$$Q_T(\theta) = T^{-1} \sum_{t=1}^{T} f(v_t, \theta)' W_T T^{-1} \sum_{t=1}^{T} f(v_t, \theta),$$

W_T is known as the the weighting matrix and is chosen to satisfy

- W_T is positive semi-definite (psd),
- $W_T \xrightarrow{p} W$, a pd matrix of constants.

Note:

- W_T is positive semi-definite (psd) $\Rightarrow$

$$Q_T(\theta) \geq 0$$

$$Q_T(\hat{\theta}_{GMM}) = 0 \text{ if } T^{-1}\sum_{i=1}^{T} f(v_t, \hat{\theta}_{GMM}) = 0$$

- $W_T \xrightarrow{p} W$ (pd) $\Rightarrow$

$$Q_T(\hat{\theta}_{GMM}) = 0 \textit{ iff } T^{-1}\sum_{i=1}^{T} f(v_t, \hat{\theta}_{GMM}) = 0 \text{ in the limit as } T \to \infty$$

Comparison of GMM to Method of Moments (MM)

$\hat{\theta}_{MM}$ is solution to $T^{-1}\sum_{t=1}^{T} f(v_t, \hat{\theta}_{MM}) = 0.$

MM only works in general if number of moments, q say, equals number of parameters, p, because if $q > p$ then no solution (even though holds in population at θ_0) due to sampling variation.

GMM works if $q \geq p$ for if:

- $q = p$ then $\hat{\theta}_{GMM} = \hat{\theta}_{MM}$.
- $q > p$ then $\hat{\theta}_{GMM}$ is value of θ that is closest to solving sample moments.

This is sense in which GMM *generalizes* MM.

It can be shown that under certain conditions:

- $\hat{\theta}_{GMM} \xrightarrow{p} \theta_0$
- $T^{1/2}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{d} N(0, V_{GMM})$ (see next slide)

There is a large array of GMM-based inference procedures available and the method is widely applied in empirical analysis.

GMM

By manipulating the FOC fo GMM estimation, it can be shown that

$$T^{1/2}(\hat{\theta}_{GMM} - \theta_0) = M_T n_T + \xi_T$$

where

- large sample behaviour is determined by $M_T n_T$
- M_T is random matrix: WLLN + Slutsky $\Rightarrow M_T \xrightarrow{p} M$, constant
- n_T is random vector: CLT $\Rightarrow n_T \xrightarrow{d} N(0, \Omega)$, where Ω pd, constant

So $V_{GMM} = M \Omega M'$ and M depends on the weighting matrix, the Jacobian matrix (derivative of the sample moment wrt θ and Ω is (LR) variance of $f(v_t, \theta_0)$.

Further reading on GMM

Material on GMM is non-examinable but FYI:

- see Hall (1993,2015) papers on BB

- Greene: Chapters 13.4-13.6