Topic 11 Header Slide

- Topic 11. Asymptotic and large sample results

  Asymptotic results: refer to theoretical results (about the probabilistic behaviour of our estimator) that hold only in the limit as S passes to ∞.

  Large sample results: refer to asymptotic results that are thought to hold approximately for sufficiently large (albeit finite) S.

Estimator	Notation
1.	$\hat{\beta}_{OLS}$
2.	$\hat{\beta}_{LAD}$
3.	$\hat{\beta}_{Lstar}$
4.	$\hat{\beta}_{GMM}$
$5a$.	$\hat{\beta}_{IGLS}$
$5b$.	$\hat{\beta}_{FGLS}$
6.	$\hat{\beta}_{MLE}$

For any method, generically denoted $\hat{\theta}_{method}$, we define

$$SEV(\hat{\theta}_{method}) = \hat{\theta}_{method} - \theta^{true}.$$

Common form of SEV for analytic methods

Consider analytic methods for estimation (i.e., whereby the objective function is twice continuously differentiable). We summarise the common structure/form of the SEV for such estimators:

- **Summary 1.** For analytic methods that are linear in y and ε^{true}, there exists a $k \times k$ matrix, B_s, and a $k \times 1$ vector, a_s, such that

$$SEV(\hat{\beta}_{method}) = \hat{\beta}_{method} - \beta^{true} = \left(\sum_{s=1}^{S} B_s\right)^{-1} \sum_{s=1}^{S} a_s.$$

  Example: We saw, for OLS, the definitions: $B_s = x_s x_s'$, and $a_s = x_s \varepsilon^{true}$.

- **Summary 2.** For analytic methods that are non-linear, there exists a $k \times k$ matrix, B_s, and a $k \times 1$ vector, a_s, such that

$$SEV(\hat{\beta}_{method}) = \hat{\beta}_{method} - \beta^{true} \approx \left(\sum_{s=1}^{S} B_s\right)^{-1} \sum_{s=1}^{S} a_s,$$

  where the approximation ("$\approx$") is reasonable for sufficiently large S, and the approximation in fact becomes exact in the limit as S passes to ∞.
  Example: We will see, for MLE in the general linear/non-linear case, the definitions: $B_s = -\ell_s^{\beta\beta'}(\beta^{true})$, and $a_s = \ell_s^{\beta}(\beta^{true})$, where ℓ_s^{β} denotes the score contribution (vector) by the s-th observation and $\ell_s^{\beta\beta'}$ denotes the corresponding second-order derivative (matrix).

Sample averages and normalised sample averages (1 of 2)

In each case above (linear/non-linear), we consider algebraic structures involving sample averages:

- **Summary 1.** For analytic methods that are linear in y and ε^{true}, there exists a $k \times k$ matrix, B_s, and a $k \times 1$ vector, a_s, such that

$$SEV(\hat{\beta}_{method}) = \hat{\beta}_{method} - \beta^{true} = \left(\frac{1}{S} \sum_{s=1}^{S} B_s \right)^{-1} \frac{1}{S} \sum_{s=1}^{S} a_s.$$

  Example: We saw, for OLS, the definitions: $B_s = x_s x_s'$, and $a_s = x_s \varepsilon^{true}$.

Above, by scaling throughout by $(1/S)$, we compute sample averages in both the inverse and non-inverse term of the SEV. This formulation of the SEV will be extremely useful to us for consistency proofs (convergence in probability).

In contrast, for asymptotic normality proofs (convergence in distribution), it is useful to consider the SEV in terms of normalised sample averages as follows:

$$\sqrt{S} SEV(\hat{\beta}_{method}) = \sqrt{S}(\hat{\beta}_{method} - \beta^{true}) = \left(\frac{1}{S} \sum_{s=1}^{S} B_s \right)^{-1} \frac{1}{\sqrt{S}} \sum_{s=1}^{S} a_s.$$

Sample averages and normalised sample averages (2 of 2)

- **Summary 2.** For analytic methods that are non-linear, there exists a $k \times k$ matrix, B_s, and a $k \times 1$ vector, a_s, such that

$$SEV(\hat{\beta}_{method}) = \hat{\beta}_{method} - \beta^{true} \approx \left(\frac{1}{S} \sum_{s=1}^{S} B_s \right)^{-1} \frac{1}{S} \sum_{s=1}^{S} a_s,$$

  where the approximation ("$\approx$") is reasonable for sufficiently large S, and the approximation in fact becomes exact in the limit as S passes to ∞.
  Example: We will see, for MLE in the general linear/non-linear case, the definitions: $B_s = -\ell_s^{\beta\beta'}(\beta^{true})$, and $a_s = \ell_s^{\beta}(\beta^{true})$, where ℓ_s^{β} denotes the score contribution (vector) by the s-th observation and $\ell_s^{\beta\beta'}$ denotes the corresponding second-order derivative (matrix).

Above, by scaling throughout by $(1/S)$, we compute sample averages in both the inverse and non-inverse term of the approximate SEV. This formulation of the approximate SEV will be extremely useful to us for consistency proofs (convergence in probability).

In contrast, for asymptotic normality proofs (convergence in distribution), it is useful to consider the approximate SEV in terms of normalised sample averages as follows:

$$\sqrt{S}(SEV(\hat{\beta}_{method})) = \sqrt{S}(\hat{\beta}_{method} - \beta^{true}) \approx \left(\frac{1}{S} \sum_{s=1}^{S} B_s \right)^{-1} \frac{1}{\sqrt{S}} \sum_{s=1}^{S} a_s.$$

Building blocks of the SEV

Our interest, broadly speaking, is in analysing what happens to $SEV(\hat{\beta}_{method})$ as $S \to \infty$. We will do so by considering each of the following sample averages or normalised sample averages:

$$\frac{1}{S}\sum_{s=1}^{S} a_s,$$

$$\frac{1}{S}\sum_{s=1}^{S} B_s,$$

$$\frac{1}{\sqrt{S}}\sum_{s=1}^{S} a_s,$$

$$\left(\frac{1}{S}\sum_{s=1}^{S} B_s\right)^{-1},$$

$$\left(\frac{1}{S}\sum_{s=1}^{S} B_s\right)^{-1} \frac{1}{S}\sum_{s=1}^{S} a_s = SEV(\hat{\beta}_{method}),$$

$$\left(\frac{1}{S}\sum_{s=1}^{S} B_s\right)^{-1} \frac{1}{\sqrt{S}}\sum_{s=1}^{S} a_s = \sqrt{S}(SEV(\hat{\beta}_{method})).$$

Extremely important remarks about the previous slide

- **Remark 1.** Recall that $\frac{1}{S}\sum_{s=1}^{S}(\cdot)$ is the first sample moment (or sample average).
- **Remark 2.** We define $\frac{1}{\sqrt{S}}\sum_{s=1}^{S}(\cdot)$ to be the first normalised sample moment (or normalised sample average).
- **Remark 3.** Inverses of matrices (unless singular) and products of matrices (unless undefined) are examples of continuous functions (or mappings) of their arguments.
- **Remark 4.** We will evaluate the behaviour of first sample moments using laws of large numbers (LLNs).
- **Remark 5.** We will evaluate the behaviour of first normalised sample moments using the central limit theorem (CLT).
- **Remark 6.** We will evaluate the behaviour of continuous functions (or mappings) of sample averages using Slutsky's theorem (i.e., our first continuous mapping theorem, CMT1).
- **Remark 7.** We will evaluate the behaviour of continuous functions (or mappings) of sample averages and normalised sample averages using Cramér's theorem (i.e., our second continuous mapping theorem, CMT2).

Specific SEVs for various analytic methods

- For method 1 (OLS) and method 4 (GMM with $A2linear$), we have

$$SEV(\hat{\beta}_{OLS}) = \hat{\beta}_{OLS} - \beta^{true} = (X'X)^{-1}X'\varepsilon^{true} = \left(\sum_{s=1}^{S} x_s x_s'\right)^{-1} \sum_{s=1}^{S} x_s \varepsilon_s^{true}.$$

- Method 2 (LAD) is not analytic; and the SEV for method 3 (Lstar) is not useful to consider.

- For method 5a (IGLS), suppose (for convenience) that we have $A4\Omega$ with a diagonal Ω. Then, we have

$$\begin{aligned} SEV(\hat{\beta}_{IGLS}) &= \hat{\beta}_{IGLS} - \beta^{true} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon^{true} \\ &= \left(\sum_{s=1}^{S} x_s \omega_{ss} x_s'\right)^{-1} \sum_{s=1}^{S} x_s \omega_{ss} \varepsilon_s^{true}, \end{aligned}$$

  where weight $\omega_{st} = \left[\Omega^{-1}\right]_{st}$ is the (s,t)-th element of Ω^{-1}, for $s, t = 1, ..., S$.

  The exact weights, ω_{st}, are unimportant; what is important is that the SEV of the IGLS estimator can be expressed in the same common form as other analytic estimators.

- We consider method 6 (MLE) in a lot of detail in the next slides.

Specific SEV for MLE (whether linear/non-linear in y)

Recall that for a parametric estimation problem for p-dimensional parameter vector θ^{true}, in the generalised LRM, via maximum likelihood estimation, we impose the assumptions $A1, A2linear, \geq A3Rmi, A4\Omega.independent$, and some $A5specific$.

In the previous sentence, $A4\Omega.independent$ refers to the $A4\Omega$ assumption with the addition of independence imposed across the s dimension (so that Ω is necessarily diagonal). Recall that this assumption makes MLE more manageable since we can thereby obtain the overall likelihood as the product of the marginal contributions to the overall likelihood by each observation.

Under the given specification, and denoting $\{y, X\}$ as $data$, we have

$$\hat{\theta}_{MLE} = \arg\max \ell(\theta; data) = \arg\max \sum_{s=1}^{S} \ell_s(\theta; data),$$

where $\ell_s(\theta; data) = \log f_s(y_s|X;\theta)$ for $s = 1, ..., S$, are the marginal contributions of each observation to the overall log likelihood.

We need to maximise an objective function but FOCs/SOCs can only be defined if the likelihood function is twice continuously differentiable. In the absence of twice continuous differentiability – e.g., under $A5LAD$ or $A5LDE$ – no analytic solution to the maximisation problem exists.

Let us suppose, for the moment, that the likelihood does admit a continuous second derivative. Now consider the maximisation procedure as outlined in the following slides.

FOC and SOC for MLE (whether linear/non-linear in y)

Under twice continuous differentiability of the likelihood function, and under the given model specification (on the previous slide), the ML estimator is (at least implicitly) defined by the following first and second order conditions (FOCs/SOCs):

FOC:

$$\frac{\partial \ell(\theta; data)}{\partial \theta}\Big|_{\theta=\hat{\theta}_{MLE}} = \sum_{s=1}^{S} \ell_s^{\theta}(\theta; data)\Big|_{\theta=\hat{\theta}_{MLE}} = \sum_{s=1}^{S} \frac{\partial \log f_s(y_s|X;\theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}_{MLE}} = 0.$$

SOC:

$$\frac{\partial^2 \ell(\theta; data)}{\partial \theta \partial \theta'}\Big|_{\theta=\hat{\theta}_{MLE}} = \sum_{s=1}^{S} \ell_s^{\theta\theta'}(\theta; data)\Big|_{\theta=\hat{\theta}_{MLE}} = \sum_{s=1}^{S} \frac{\partial \log f_s(y_s|X;\theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}_{MLE}}$$

is negative definite.

Note above that first order derivative of the log of the likelihood (also called the "score function") is a $p \times 1$ vector; and the second order derivative of the log of the likelihood is a $p \times p$ matrix.

Approximate SEV for MLE (whether linear/non-linear in y)

Focussing on the FOCs for a moment, we had under twice continuous differentiability of the likelihood function and the given model specification, that the ML estimator is defined by the following system of p equations in p unknowns:

FOC:

$$\sum_{s=1}^{S} \ell_s^{\theta}(\theta; data)\Big|_{\theta=\hat{\theta}_{MLE}} = 0.$$

Now, under $A5Gaussian$ for example, we can solve explicitly for $\hat{\theta}_{MLE}$. But what if we have a different $A5specific$, which although twice continuously differentiable, does not admit a closed-form expression for the ML estimator? In other words, what if the score function is non-linear in θ? (Ans: We find a linear approximation to the score function at θ^{true} and set that to zero!)

Consider the first order Taylor expansion of $\sum_{s=1}^{S} \ell_s^{\theta}(\theta; data)$ at θ^{true} given by

$$LHS = \sum_{s=1}^{S} \ell_s^{\theta}(\hat{\theta}_{MLE}; data) \approx \sum_{s=1}^{S} \ell_s^{\theta}(\theta^{true}; data) + \sum_{s=1}^{S} \ell_s^{\theta\theta'}(\theta^{true}; data)(\hat{\theta}_{MLE} - \theta^{true}) = RHS$$

The simple intuition is that since we cannot directly set LHS to zero and solve, we set a first order approximation of the LHS – i.e., the RHS – to zero and solve that instead.

Approximate SEV for MLE (whether linear/non-linear in y)

Continuing the analysis on the previous slide, we have that

$$\sum_{s=1}^{S} \ell_s^{\theta}(\theta^{true}; data) + \sum_{s=1}^{S} \ell_s^{\theta\theta'}(\theta^{true}; data)(\hat{\theta}_{MLE} - \theta^{true}) \approx 0,$$

so that by rearranging, we obtain

$$\left(\hat{\theta}_{MLE} - \theta^{true}\right) \approx - \left(\sum_{s=1}^{S} \ell_s^{\theta\theta'}(\theta^{true}; data) \right)^{-1} \sum_{s=1}^{S} \ell_s^{\theta}(\theta^{true}; data),$$

as the approximate SEV for ML estimator where the approximation is considered reasonable for sufficiently large sample size, S.

Asymptotic results for the OLS estimator in the ANLRM – scenario 1

Let us end this topic on asymptotics by summarising the asymptotic results available for the OLS estimator under the ANLRM. Notice that we are able to weaken our exogeneity assumption to "$\geq A3Rsru$" when considering results that are available only asymptotically as $S \to \infty$ (i.e., results that are not necessarily "exact").

Suppose we have ANLRM.$A4GM(iid)$ where $A1, A2linear, \geq A3Rsru, A4GM(iid)$. Then, we can/will prove (by LLNs, CLT and CMT2) that

$$\sqrt{S}(SEV(\hat{\beta}_{OLS})) = \sqrt{S}(\hat{\beta}_{OLS} - \beta^{true})|X \xrightarrow{d} N\left(0, \sigma_\varepsilon^2 \left(p \lim_{S\to\infty} \frac{X'X}{S}\right)^{-1}\right)$$

as $S \to \infty$.

Alternatively, for inferential purposes, the previous statement is taken as justification to say

$$\hat{\beta}_{OLS}|X \overset{approx}{\sim} N\left(\beta^{true}, \sigma_\varepsilon^2 (X'X)^{-1}\right)$$

for sufficiently large S.

Asymptotic results for the OLS estimator in the ANLRM – scenario 2

Suppose we have ANLRM.$A4\Omega$ where $A1, A2linear, \geq A3Rsru, A4\Omega$. Then, we can/will prove (by LLNs, CLT and CMT2) that

$$\sqrt{S}(SEV(\hat{\beta}_{OLS})) = \sqrt{S}(\hat{\beta}_{OLS} - \beta^{true})|X \xrightarrow{d} N\left(0, c^2 Q\right)$$

where

$$Q = \left(p \lim_{S\to\infty} \frac{X'X}{S}\right)^{-1} \left(p \lim_{S\to\infty} \frac{X'\Omega X}{S}\right) \left(p \lim_{S\to\infty} \frac{X'X}{S}\right)^{-1}$$

as $S \to \infty$.

Alternatively, for inferential purposes, the previous statement is taken as justification to say

$$\hat{\beta}_{OLS}|X \overset{approx}{\sim} N\left(\beta^{true}, c^2(X'X)^{-1}X'\Omega X(X'X)^{-1}\right)$$

for sufficiently large S.

Appreciating the distinction between different types of results

- Suppose we have NLRM.$A4GM(iid)$ where $A1, A2linear, \geq\ A3Rmi, A4GM(iid)$, and $A5Gaussian$ hold. Then, we can prove (due to preservation of multivariate Gaussianity) that

$$\hat{\beta}_{OLS}|X{\sim}N\left(\beta^{true}, \sigma_{\varepsilon}^{2}(X'X)^{-1}\right),$$

  which is true for any S (even finite). This is an exact or finite-sample result true for any S.

- Suppose we have ANLRM.$A4GM(iid)$ where $A1, A2linear, \geq\ A3Rsru, A4GM(iid)$. Then, we can/will prove (by LLNs, CLT and CMT2) that

$$\sqrt{S}(SEV(\hat{\beta}_{OLS})) = \sqrt{S}(\hat{\beta}_{OLS} - \beta^{true})|X \xrightarrow{d} N\left(0, \sigma_{\varepsilon}^{2}\left(p\lim_{S\to\infty}\frac{X'X}{S}\right)^{-1}\right)$$

  as $S \to \infty$. This is an asymptotically valid result true only as $S \to \infty$.

- Alternatively, for inferential purposes, the previous statement is taken as justification to say

$$\hat{\beta}_{OLS}|X \overset{approx}{\sim} N\left(\beta^{true}, \sigma_{\varepsilon}^{2}(X'X)^{-1}\right)$$

  for sufficiently large S. This is an approximate result that is appropriate for large S.