Steven M Bergantino
14.382
Handout #6

OLS: LARGE SAMPLE RESULTS

Assumptions

(A0) $y = X\beta + \varepsilon$	Linear Functional Form
(A1) $X'X/n \xrightarrow{p} M$, Pos. Def.	Asymptotic Identification
(A2) $X'\varepsilon/n \xrightarrow{p} 0$	Asymptotic Orthogonality
(A3) $X'\varepsilon/\sqrt{n} \xrightarrow{d} N(0, V)$, $V = \text{plim}(X'\Omega X/n)$	Asymptotic Normality
(A4) $\Omega = \sigma^2 I$ and $\varepsilon'\varepsilon/n \xrightarrow{p} \sigma^2$	Sphericality

where $\Omega_{ij} = E(\varepsilon_i \varepsilon_j | X_i, X_j)$ $\forall i,j$ and where we continue to assume that rank$(X) = k$ in all samples. If $\hat{\beta}$ is the OLS estimator of β and $s^2 = \hat{\varepsilon}'\hat{\varepsilon}/(n-k)$ then we can establish the following large sample results.

(T1) Theorem: Under (A0) - (A2), $\hat{\beta} \xrightarrow{p} \beta$

(T2) Theorem: Under (A0) - (A3)

(1) $\hat{\beta} \xrightarrow{p} \beta$

(2) $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, M^{-1} V M^{-1})$

(3) $\dfrac{C'(\hat{\beta} - \beta)}{SE(C'\hat{\beta})} \xrightarrow{d} N(0,1)$

where $SE(C'\hat{\beta}) = \left[C'(X'X)^{-1} X'\hat{\Omega} X (X'X)^{-1} C\right]^{1/2}$ for any $\hat{\Omega}$ w/ $\frac{X'\hat{\Omega}X}{n} \xrightarrow{p} V$

(4) $[h(\hat{\beta}) - h(\beta)]' \left[H(\hat{\beta})(X'X)^{-1} X'\hat{\Omega} X (X'X)^{-1} H(\hat{\beta})'\right]^{-1} [h(\hat{\beta}) - h(\beta)] \xrightarrow{d} \chi^2_J$

for any $J \times 1$ function $h(\cdot)$ which is continuously differentiable at β and which has derivative matrix $H(\beta) \equiv \partial h(\beta)/\partial \beta'$ $(J \times K)$ with full row rank $J \leq K$ in a neighborhood around β.

(C2a) <u>Corollary</u>: An asymptotically valid test of the hypotheses

(1) $H_0: C'\beta = a$ vs. $H_1: C'\beta \neq a$ is reject H_0 if $|T| \geq Z_{\alpha/2}$
(2) $H_0: C'\beta \leq a$ vs. $H_1: C'\beta \geq a$ is reject H_0 if $T \geq Z_\alpha$
(3) $H_0: C'\beta \geq a$ vs. $H_1: C'\beta \leq a$ is reject H_0 if $T \leq -Z_\alpha$

where $T = (c'\hat{\beta} - a)/SE(C'\hat{\beta})$ and $\mathbb{P}(N(0,1) \geq Z_\alpha) = \alpha$

<u>Note:</u> This includes the standard t-test of $\beta_k = 0$ (with robust standard errors) by letting C be the kth column of I_k.

(C2b) <u>Corollary</u>: An asymptotically valid test of the hypotheses

$H_0: h(\beta) = 0$ vs. $H_1: h(\beta) \neq 0$ is Reject H_0 if $W \geq C_\alpha$

where $W = h(\hat{\beta})'[H(\hat{\beta})(x'x)^{-1}x'\hat{\Omega}x(x'x)^{-1}H(\hat{\beta})']^{-1}h(\hat{\beta})$ and $\mathbb{P}(\chi^2_J \geq C_\alpha) = \alpha$

<u>Note:</u> This includes tests of linear restrictions $R\beta = c$, using the usual Wald statistic w/ $S^2(x'x)^{-1}$ replaced by $(x'x)^{-1}x'\hat{\Omega}x(x'x)^{-1}$.
! <u>However</u>, the other forms of this test based on loss of fit (ie $\hat{\varepsilon}'\hat{\varepsilon}$ and R^2 forms of test) are <u>not</u> valid because they are numerically equivalent to the $S^2(x'x)^{-1}$ form of the Wald statistic, which is innappropriate for general Ω.

(T3) <u>Theorem:</u> Under (A0) - (A5)

(1) $\hat{\beta} \xrightarrow{p} \beta$
(2) $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 M^{-1})$
(3) $S^2 \xrightarrow{p} \sigma^2$
(4) $C'(\hat{\beta} - \beta)/SE(C'\hat{\beta}) \xrightarrow{d} N(0,1)$ where $SE(C'\hat{\beta}) = [S^2 C'(x'x)^{-1}C]^{1/2}$

(5) $[h(\hat{\beta}) - h(\beta)]'[S^2 H(\hat{\beta})(x'x)^{-1}H(\hat{\beta})']^{-1}[h(\hat{\beta}) - h(\beta)] \xrightarrow{d} \chi^2_J$

For any $J \times 1$ function $h(\cdot)$, continuously differentiable at β and with $H(\beta) \equiv \partial h(\beta)/\partial \beta'$ $(J \times K)$ having full row rank $J \leq K$ in a neigborhood around β.

(C3a) <u>Corollary</u>: Same as (C2a) but w/ $SE(C'\hat{\beta})$ defined as in (T3)

(C3b) <u>Corollary</u>: Same as (C2b) but w/ $S^2(X'X)^{-1}$ replacing $(X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1}$

<u>Note</u>: This includes tests of linear restrictions $R\beta = c$. Also, the loss of fit based forms of the test are now also asymptotically valid.

SMALL SAMPLE VS LARGE SAMPLE

Small Sample Assumptions		Large Sample Assumptions
(SS0) $Y = X\beta + \varepsilon$	linear f^n form	(LS0) $Y = X\beta + \varepsilon$
(SS1) $\text{rank}(X) = K$	identification	(LS1) $X'X/n \xrightarrow{p} M$, PD and $\text{rank}(X) = K$
(SS2) $E(\varepsilon \mid X) = 0$	(asy) orthogonality	(LS2) $X'\varepsilon/n \xrightarrow{p} 0$
(SS3) $\varepsilon \sim N(0, \Omega) \mid X$	(asy) normality	(LS3) $X'\varepsilon/\sqrt{n} \xrightarrow{d} N(0, V)$, $V = \text{plim}(X'\Omega X/n)$
(SS4) $\Omega = \sigma^2 I$	sphericality	(LS4) $\Omega = \sigma^2 I$ and $\varepsilon'\varepsilon/n \xrightarrow{p} \sigma^2$

Small Sample Results	Large Sample Results
(SS0) – (SS2)	(LS0) – (LS2)
• $E(\hat{\beta}) = \beta$	• $\hat{\beta} \xrightarrow{p} \beta$
(SS0) – (SS3)	(LS0) – (LS3)
• $\hat{\beta} \sim N(\beta, \underbrace{(X'X)^{-1}X'\Omega X(X'X)^{-1}}_{V}) \mid X$ • There are no small sample test statistics. The White and Newey-West estimators of V that we will discuss later rely on asymptotic results and so are only valid in large samples.	• $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, M^{-1}VM^{-1})$ $\Rightarrow \hat{\beta} \overset{a}{\sim} N(\beta, (X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1})$ • $C'(\hat{\beta} - \beta)/SE(C'\hat{\beta}) \xrightarrow{d} N(0,1)$ $\Rightarrow$ asymptotically valid Z-tests (or t-tests) • $[h(\hat{\beta}) - h(\beta)]'[H(\hat{\beta})(X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1}H(\hat{\beta})']^{-1}[h(\hat{\beta}) - h(\beta)] \xrightarrow{d} \chi^2_J$ $\Rightarrow$ asymptotically valid Chi-squared tests (or F-tests) of <u>linear</u> and <u>nonlinear</u> restrictions.
(SS0) – (SS2), (SS4)	
• GMT $\Rightarrow$ $\hat{\beta}$ is BLUE <u>Note</u>: There is no large sample equivalent for (LS0) – (LS2), (LS4)	<u>Note</u>: requires $\hat{\Omega}$ s.t. $X'\hat{\Omega}X/n \xrightarrow{p} V$ (ie: $\hat{\Omega}$ s.t. $\hat{\Omega} \xrightarrow{p} \Omega$ (a consistent est. of Ω) and Therefore $\frac{X'\hat{\Omega}X}{n} \xrightarrow{p} \frac{X'\Omega X}{n} \xrightarrow{p} V$)

(SS0) – (SS4)	(LS0) – (LS4)
• $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1}) \mid X$	• $\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} N(0, \sigma^2 M^{-1})$
• $(n-k)S^2/\sigma^2 \sim \chi^2_{n-k}$ ind of $\hat{\beta}$ $\Rightarrow ES^2 = \sigma^2$	• $S^2 \xrightarrow{p} \sigma^2$ $\Rightarrow \hat{\beta} \overset{a}{\sim} N(\beta, S^2(X'X)^{-1})$
• $C'(\hat{\beta}-\beta)/SE(C'\hat{\beta}) \sim t_{n-k}$ $\Rightarrow$ Exact small sample t-tests	• $C'(\hat{\beta}-\beta)/SE(C'\hat{\beta}) \xrightarrow{d} N(0,1)$ $\Rightarrow$ asymptotically valid Z-tests (or t-tests)
• $(R\hat{\beta}-R\beta)'[S^2R(X'X)^{-1}R']^{-1}(R\hat{\beta}-R\beta) \sim F(J, n-k)$ $\Rightarrow$ Exact small sample F-tests of <u>linear</u> restrictions	• $[h(\hat{\beta})-h(\beta)]'[S^2H(\beta)(X'X)^{-1}H(\beta)']^{-1}[h(\hat{\beta})-h(\beta)]$ $\xrightarrow{d} \chi^2_J$ $\Rightarrow$ asymptotically <u>valid Chi-Squared</u> tests (or F-tests) of <u>linear</u> and <u>nonlinear</u> restrictions
• $\hat{\beta}$ is MLE of $\beta \Rightarrow \hat{\beta}$ is BUE	
	• By (T18) HO3, MLE is asymptotically efficient. Thus, <u>in order for $\hat{\beta}$ to be asymptotically efficient</u> we would have to make the additional assumption that $\varepsilon_i \sim N(0, \sigma^2)$.

<u>Remarks:</u>

• The Wald and F-statistics we derived in HO5 for testing $R\beta = C$ satisfy the convergence results $W \xrightarrow{d} \chi^2_J$ and $FJ \xrightarrow{d} \chi^2_J$ under H_0 and (LS0) – (LS4). Thus, we can do asymptotically valid Chi-Squared tests of linear restrictions. Often, however, conservatism toward rejecting H_0 leads people to use F-tests anyway. This is OK. For nonlinear tests, however, people always use the Chi-Squared tables, presumably because there is no set of assumptions under which they have small sample F-distributions. Similarly, people often use t-tests rather than Z-tests even when operating under the large sample assumptions.

• Asymptotic efficiency of OLS requires an addition assumption of normally distributed errors. Since making such an assumption doesn't change the asymptotic distribution of $\hat{\beta}$ ($\hat{\beta}_{ML} = \hat{\beta}_{OLS}$ in this case) people often implicitly or explicitly make this assumption (eg GLS). This is fine unless one has strong evidence that the errors are very far from being normal (eg are skewed or have fat tails) in which case OLS is most likely not equivalent to doing true MLE, making claims of asymptotic efficiency suspect.

- Whitney's handout gives more primitive assumptions under which assumptions (LS1) – (LS4) hold and, hence, under which the stated theorems hold. No point in copying his (big-ugly-deep) proofs so I refer you to him and his handout for generalizations of the results, except for the case of beautifully behaved iid data which I will cover in Appendix 2.

APPENDIX 1: (PROOFS)

It should be clear that all theorems and corollaries in this handout follow more or less directly from theorem (T2). I will, therefore, do only the proof of this theorem and result (3) of (T3).

Proof (of (T2)):

(1) $\hat{\beta} = (X'X)^{-1}X'Y$

$= (X'X)^{-1}X'(X\beta + \varepsilon) \leftarrow$ by (A0)

$= \beta + (X'X/n)^{-1}(X'\varepsilon/n) \xrightarrow{p} \beta + M^{-1}0 = \beta + 0 = \beta$ by (C3b) (b) of HO3

$(X'X/n)^{-1} \xrightarrow{p} M^{-1}$ by (C3b) (c) of HO3 given (A1)

$(X'\varepsilon/n) \xrightarrow{p} 0 \leftarrow$ by (A2)

(2) $\sqrt{n}(\hat{\beta} - \beta) = (X'X/n)^{-1}(X'\varepsilon/\sqrt{n}) \xrightarrow{d} M^{-1}N(0,V) = N(0, M^{-1}VM^{-1})$

by (C5b) (b) of HO3 given (A3) and given $(X'X/n)^{-1} \xrightarrow{p} M^{-1}$ (shown in (1)).

(3) $C'\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} C'N(0, M^{-1}VM^{-1}) = N(0, C'M^{-1}VM^{-1}C)$

by (C5b) (b) of HO3 and result (2)

$C'\left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\hat{\Omega}X}{n}\right)\left(\frac{X'X}{n}\right)^{-1}C \xrightarrow{p} C'M^{-1}VM^{-1}C$ by (C3b) (b) of HO3

for $\hat{\Omega}$ such that $\left(\frac{X'\hat{\Omega}X}{n}\right) \xrightarrow{p} V$. Thus, by Slutsky's Theorem

((T3) of HO3) we have

$$\left[C'\left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\hat{\Omega}X}{n}\right)\left(\frac{X'X}{n}\right)^{-1}C\right]^{1/2} \xrightarrow{p} \left[C'M^{-1}VM^{-1}C\right]^{1/2}$$ and so

$$\frac{C'(\hat{\beta}-\beta)}{SE(C'\hat{\beta})} = \frac{\sqrt{n}\,C'(\hat{\beta}-\beta)}{\left[C'\left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\hat{\Omega}X}{n}\right)\left(\frac{X'X}{n}\right)^{-1}C\right]^{1/2}} \xrightarrow{d} \frac{N(0, C'M^{-1}VM^{-1}C)}{\left[C'M^{-1}VM^{-1}C\right]^{1/2}} = N(0,1)$$

by (C5a) (c) of HO3.

(4) $\sqrt{n}(h(\hat{\beta})-h(\beta)) \xrightarrow{d} N(0, H(\beta)M^{-1}VM^{-1}H(\beta)')$ by the Delta Method ((T16) of HO3)

$H(\hat{\beta}) \xrightarrow{p} H(\beta)$ by Slutsky's Theorem ((T3) of HO3) and result (1)

$$\Rightarrow H(\hat{\beta})\left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\hat{\Omega}X}{n}\right)\left(\frac{X'X}{n}\right)^{-1}H(\hat{\beta})' \xrightarrow{p} H(\beta)M^{-1}VM^{-1}H(\beta)'$$ by (C3b) (L)

of HO3, and, since by result (1) this is PD w/ probability approaching 1 as $n \to \infty$ ($\hat{\beta} \xrightarrow{p} \beta \Rightarrow H(\hat{\beta})$ full row rank w/ prob $\to 1$), we have

$$\left[H(\hat{\beta})\left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\hat{\Omega}X}{n}\right)\left(\frac{X'X}{n}\right)^{-1}H(\hat{\beta})'\right]^{-1/2} \xrightarrow{p} \left[H(\beta)M^{-1}VM^{-1}H(\beta)'\right]^{-1/2}$$

by Slutsky's Theorem. Thus, by (C5b) (b) of HO3

$$\underbrace{\left[H(\hat{\beta})\left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\hat{\Omega}X}{n}\right)\left(\frac{X'X}{n}\right)^{-1}H(\hat{\beta})'\right]^{-1/2}\sqrt{n}\left[h(\hat{\beta})-h(\beta)\right]}_{Z_n}$$

$$\xrightarrow{d} \left[H(\beta)M^{-1}VM^{-1}H(\beta)'\right]^{-1/2} N(0, H(\beta)M^{-1}VM^{-1}H(\beta)') = \overbrace{N(0, I_J)}^{Z}$$

and (noticing that the n's all cancel) we can apply the Continuous Mapping Theorem ((T5) of HO3) to $Z_n'Z_n$ to get:

$$[h(\hat{\beta})-h(\beta)]'\left[H(\hat{\beta})(X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1}H(\hat{\beta})'\right]^{-1}[h(\hat{\beta})-h(\beta)] = Z_n'Z_n$$

$$\xrightarrow{d} Z'Z \sim \chi^2_J \;//$$

Proof (of (T3) (3))

$$S^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k} = \frac{(QY)'(QY)}{n-k} = \frac{[Q(X\beta+\varepsilon)]'[Q(X\beta+\varepsilon)]}{n-k} = \frac{(Q\varepsilon)'(Q\varepsilon)}{n-k}$$

$$= \frac{\varepsilon' Q \varepsilon}{n-k} = \frac{n}{n-k}\,\frac{\varepsilon'[I - X(X'X)^{-1}X']\varepsilon}{n}$$

$$= \underbrace{\frac{n}{n-k}}_{\to 1}\left[\underbrace{\frac{\varepsilon'\varepsilon}{n}}_{\xrightarrow{p}\sigma^2 \text{ by (A4)}} - \underbrace{\frac{\varepsilon'X}{n}}_{\xrightarrow{p} 0}\underbrace{\left(\frac{X'X}{n}\right)^{-1}}_{\xrightarrow{p} M^{-1}}\underbrace{\frac{X'\varepsilon}{n}}_{\xrightarrow{p} 0}\right] \xrightarrow{p} \sigma^2 \quad //$$

APPENDIX 2: THE I.I.D. CASE

(A0) $Y = X\beta + \varepsilon$ and $\{X_i, \varepsilon_i\}$ iid w/ finite fourth moments
(A1') $E X_i X_i'$ is PD
(A2') $E X_i \varepsilon_i = 0$
(A3') $E(\varepsilon_i^2 | X_i) = \sigma^2$

(T4) Theorem: (T3) holds under these assumptions

Proof:
We just need to show that (A1') - (A3') $\Rightarrow$ (A1) - (A4).

$$\frac{X'X}{n} = \frac{1}{n}\sum_1^n X_i X_i' \xrightarrow{p} \underbrace{E X_i X_i'}_{M} \text{ PD by WLLN} \Rightarrow \text{(A1)}$$

$$\frac{X'\varepsilon}{n} = \frac{1}{n}\sum_1^n X_i \varepsilon_i \xrightarrow{p} E X_i \varepsilon_i = 0 \text{ by WLLN} \Rightarrow \text{(A2)}$$

$$\frac{X'\varepsilon}{\sqrt{n}} = \sqrt{n}\left(\frac{1}{n}\sum_1^n X_i \varepsilon_i\right) \xrightarrow{d} N(0, \underbrace{\sigma^2 E X_i X_i'}_{V}) \text{ by Lindberg-Levy} \Rightarrow \text{(A}$$

since $X_i \varepsilon_i$ is iid w/ $E(X_i\varepsilon_i) = 0$ and $V(X_i \varepsilon_i) = \sigma^2 E(X_i X_i')$. Also,

$$\frac{X'\Omega X}{n} = \sigma^2 \frac{X'X}{n} \xrightarrow{p} \sigma^2 E X_i X_i' = V \text{ (since } \Omega = \sigma^2 I \text{ by (A3') and iid) an}$$

$$\frac{\varepsilon'\varepsilon}{n} = \frac{1}{n}\sum_1^n \varepsilon_i^2 \xrightarrow{p} E\varepsilon_i^2 = E(E(\varepsilon_i^2 | X_i)) = \sigma^2 \text{ by WLLN} \Rightarrow \text{(A4)}$$