LECTURE 8

LARGE SAMPLE PROPERTIES OF OLS, ASYMPTOTIC CONFIDENCE INTERVALS AND HYPOTHESIS TESTING

In this lecture, we discuss large sample properties of the OLS estimator for the linear regression model defined by the following assumptions:

(A1) $Y_i = X_i'\beta + U_i$.

(A2*) $E(U_i X_i) = 0$.

(A6) $\{(Y_i, X_i) : i = 1, \ldots, n\}$ iid.

In addition, the large sample properties will be derived under one or more of the following assumptions:

(A7) $E(X_i X_i')$ is a finite positive definite matrix.

(A8) $EX_{i,j}^4 < \infty$ for all $j = 1, \ldots, k$.

(A9) $EU_i^4 < \infty$.

(A10) $EU_i^2 X_i X_i'$ is positive definite.

Consistency

The estimator $\widehat{\beta}_n$ is *consistent* for β if $\widehat{\beta}_n \to_p \beta$ as $n \to \infty$. Write the OLS estimator of β as

$$\widehat{\beta}_n = \left(\sum_{i=1}^n X_i X_i' \right)^{-1} \sum_{i=1}^n X_i Y_i.$$

The following theorem gives the consistency of the OLS.

Theorem 1 *Under Assumptions (A1), (A2*), (A6) and (A7), $\widehat{\beta}_n \to_p \beta$ as $n \to \infty$.*

Proof. First, note that U_i's and $X_i U_i$'s are iid. Write

$$\widehat{\beta}_n = \beta + \left(n^{-1} \sum_{i=1}^n X_i X_i' \right)^{-1} n^{-1} \sum_{i=1}^n X_i U_i. \tag{1}$$

By the WLLN,[1]

$$\begin{aligned} n^{-1} \sum_{i=1}^n X_i U_i &\to_p E(X_1 U_1) \\ &= 0. \end{aligned}$$

[1] Let X be a random variable and define

$$\begin{aligned} X^+ &= \max(0, X), \\ X^- &= \max(0, -X), \end{aligned}$$

so that

$$X = X^+ - X^-.$$

Note that both X^+ and X^- are nonnegative random variables. When at least one of the following conditions holds: $EX^+ < \infty$ or $EX^- < \infty$, the expected value of X is given by

$$EX = EX^+ - EX^-.$$

The expectation EX is *not defined* when $EX^+ = \infty$ and $EX^- = \infty$ (thus, we prohibit $\infty - \infty$). Since

$$|X| = X^+ + X^-,$$

we have that $E|X| < \infty$ if and only if $EX^+ < \infty$ and $EX^- < \infty$. When we say that $EX = \mu$ for some μ, we therefore assume that either $EX^+ < \infty$ or $EX^- < \infty$ in order for EX to be defined. If μ is finite, it has to be the case that $EX^+ < \infty$ and $EX^- < \infty$ and, consequently, $E|X| < \infty$.

Since $E(X_1 X_1')$ is finite, the WLLN implies that

$$n^{-1}\sum_{i=1}^{n} X_i X_i' \to_p E(X_1 X_1').$$

Since $E(X_1 X_1')$ is positive definite, it follows from the Slutsky's Theorem that

$$\left(n^{-1}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \to_p (E(X_1 X_1'))^{-1}. \tag{2}$$

Hence,

$$\left(n^{-1}\sum_{i=1}^{n} X_i X_i'\right)^{-1} n^{-1}\sum_{i=1}^{n} X_i U_i \to_p 0,$$

and, therefore, $\widehat{\beta}_n \to_p \beta$. □

Asymptotic normality

In this section, we describe the asymptotic distribution of $\widehat{\beta}_n$.

Theorem 2 *Under Assumptions (A1), (A2*), (A6) - (A10),* $n^{1/2}\left(\widehat{\beta}_n - \beta\right) \to_d N(0, V)$, *where*

$$\begin{aligned} V &= Q^{-1}\Omega Q^{-1}, \\ Q &= E(X_1 X_1'), \\ \Omega &= E\left(U_1^2 X_1 X_1'\right). \end{aligned}$$

Proof. Re-write (1) as

$$n^{1/2}\left(\widehat{\beta}_n - \beta\right) = \left(n^{-1}\sum_{i=1}^{n} X_i X_i'\right)^{-1} n^{-1/2}\sum_{i=1}^{n} X_i U_i.$$

Consider $n^{-1/2}\sum_{i=1}^{n} X_i U_i$. By Assumption (A2*), $E(X_1 U_1) = 0$. Next, consider $Var(X_1 U_1) = E\left(U_1^2 X_1 X_1'\right)$. The (r, s) element of $Var(X_1 U_1)$ is $E\left(U_1^2 X_{1,r} X_{1,s}\right)$. By the Cauchy-Schwartz inequality, and due to Assumptions (A8) and (A9)

$$\begin{aligned} E\left|U_1^2 X_{1,r} X_{1,s}\right| &\le \left(EU_1^4 E\left(X_{1,r}^2 X_{1,s}^2\right)\right)^{1/2} \\ &\le \left(EU_1^4\right)^{1/2}\left(EX_{1,r}^4 EX_{1,s}^4\right)^{1/4} \\ &< \infty. \end{aligned}$$

By the CLT

$$\begin{aligned} n^{-1/2}\sum_{i=1}^{n} X_i U_i &\to_d N\left(0, E\left(U_1^2 X_1 X_1'\right)\right) \\ &= N(0, \Omega). \end{aligned} \tag{3}$$

Finally, it follows from (2), (3) and the Cramer Convergence Theorem (multivariate extension), that

$$\begin{aligned} \left(n^{-1}\sum_{i=1}^{n} X_i X_i'\right)^{-1} n^{-1/2}\sum_{i=1}^{n} X_i U_i &\to_d Q^{-1} N(0, \Omega) \\ &= N\left(0, Q^{-1}\Omega Q^{-1}\right). \end{aligned}$$

□

Remarks:

1. The assumptions of the theorem allow for the conditional variance of U_i's to depend on X_i, i.e. it is possible that the errors U_i's are *heteroskedastic*: $E\left(U_i^2|X_i\right) = \sigma^2(X_i)$ for some function $\sigma^2 : R^k \to R$.

2. The asymptotic variance-covariance matrix of $\widehat{\beta}_n$ is given by the "sandwich" formula

$$V = \left(E\left(X_1X_1'\right)\right)^{-1} E\left(U_1^2 X_1 X_1'\right)\left(E\left(X_1X_1'\right)\right)^{-1}.$$

If Assumption (A3), $E\left(U_1^2|X_1\right) = \sigma^2$, holds, then, V simplifies to the *homoskedastic* variance given by $\sigma^2\left(E\left(X_1X_1'\right)\right)^{-1}$. First, by the LIE

$$\begin{aligned} E\left(U_1^2X_1X_1'\right) &= EE\left(U_1^2X_1X_1'|X\right) \\ &= E\left(X_1X_1'E\left(U_1^2|X\right)\right) \\ &= \sigma^2 E\left(X_1X_1'\right). \end{aligned}$$

Therefore, in this case,

$$\Omega = \sigma^2 Q,$$

and

$$\begin{aligned} V &= Q^{-1}\Omega Q^{-1} \\ &= \sigma^2 Q^{-1}. \end{aligned}$$

Variance-covariance matrix estimation

Given the estimator of β, construct the fitted residuals as $\widehat{U}_i = Y_i - X_i'\widehat{\beta}_n$.Consider the following estimator of V implied by the MM principle:

$$\begin{aligned} \widehat{V}_n &= \widehat{Q}_n^{-1}\widehat{\Omega}_n\widehat{Q}_n^{-1}, \text{ where} \\ \widehat{Q}_n &= n^{-1}\sum_{i=1}^n X_iX_i', \\ \widehat{\Omega}_n &= n^{-1}\sum_{i=1}^n \widehat{U}_i^2 X_iX_i'. \end{aligned}$$

We have shown above that $\widehat{Q}_n^{-1} \to_p Q^{-1}$. Next, consider $\widehat{\Omega}_n$ and note that $\widehat{U}_i$ depends on all U_i's, $i = 1, \ldots, n$, through $\widehat{\beta}_n$, and, consequently, the WLLN cannot be applied directly to the averages containing $\widehat{U}_i$. Further, for $\widehat{\Omega}_n$, one cannot rely on properties (ii) and (iii) of the probability limits from Lecture 7, since $\widehat{\Omega}_n$ has n elements $\widehat{U}_i^2X_iX_i'$ in the sum. Write

$$\begin{aligned} \widehat{U}_i &= Y_i - X_i'\widehat{\beta}_n \\ &= U_i - X_i'\left(\widehat{\beta}_n - \beta\right). \end{aligned}$$

Therefore,

$$\begin{aligned} n^{-1}\sum_{i=1}^n \widehat{U}_i^2X_iX_i' &= n^{-1}\sum_{i=1}^n U_i^2X_iX_i' - 2R_{1,n} + R_{2,n}, \text{ where} \qquad (4) \\ R_{1,n} &= n^{-1}\sum_{i=1}^n \left(\left(\widehat{\beta}_n - \beta\right)' X_iU_i\right)X_iX_i' \\ R_{2,n} &= n^{-1}\sum_{i=1}^n \left(\left(\widehat{\beta}_n - \beta\right)' X_i\right)^2 X_iX_i'. \end{aligned}$$

Under the Assumptions specified in Theorem 2, $E\left(U_i^2 X_i X_i'\right)$ is finite, as was shown in its proof. Consequently, by the WLLN,

$$n^{-1} \sum_{i=1}^{n} U_i^2 X_i X_i' \to_p E\left(U_1^2 X_1 X_1'\right).$$

One can show that $R_{1,n}$ and $R_{2,n}$ converge in probability to zero (see the Appendix), and, consequently,

$$\widehat{V}_n \to_p V.$$

The variance-covariance matrix estimator $\widehat{V}_n = \widehat{Q}_n^{-1} \widehat{\Omega}_n \widehat{Q}_n^{-1}$ relies on the sandwich formula, and, therefore, gives consistent estimates of the asymptotic variance of the OLS in the cases of homoskedastic or heteroskedastic errors. It is often called robust, heteroskedasticity consistent or the White's estimator (it was suggested by White (1980), *Econometrica).* Many statistical software packages (Eviews, SAS, Stata) can compute standard errors using the White's estimator, however, by default they usually produce standard errors for the homoskedastic case using $s^2 \left(n^{-1} \sum_{i=1}^{n} X_i X_i'\right)^{-1}$ as an estimator (note that s^2 is now a sequence indexed by n; it is easy to show that $s^2 \to_p \sigma^2 = EU_1^2$).

Asymptotic confidence intervals

In this section we discuss asymptotic confidence intervals for the elements of β. Consider the following confidence interval for the j-th element of β:

$$CI_{n,j,1-\alpha} = \left[\widehat{\beta}_{n,j} - z_{1-\alpha/2} \sqrt{\left[\widehat{V}_n\right]_{jj} / n}, \widehat{\beta}_{n,j} + z_{1-\alpha/2} \sqrt{\left[\widehat{V}_n\right]_{jj} / n}\right],$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution, and $\left[\widehat{V}_n\right]_{jj}$ denotes the (j,j) element of the matrix $\widehat{V}_n$. We will show that $P\left(\beta_j \in CI_{n,j,1-\alpha}\right) \to 1-\alpha$ as $n \to \infty$. Since $n^{1/2}\left(\widehat{\beta}_n - \beta\right) \to_d N(0,V)$, and $\widehat{V}_n \to_p V$, it follows from the Slutsky's and Cramer Convergence Theorems that

$$\begin{aligned} \widehat{V}_n^{-1/2} n^{1/2} \left(\widehat{\beta}_n - \beta\right) &\to_d V^{-1/2} N(0,V) \\ &= N(0, I_k), \end{aligned}$$

and, consequently,

$$\frac{\sqrt{n}\left(\widehat{\beta}_{n,j} - \beta_j\right)}{\sqrt{\left[\widehat{V}_n\right]_{jj}}} \to_d N(0,1),$$

which can be also written as

$$P\left(\frac{\sqrt{n}\left(\widehat{\beta}_{n,j} - \beta_j\right)}{\sqrt{\left[\widehat{V}_n\right]_{jj}}} \le z\right) \to P(Z \le z) \text{ for all } z \in R,$$

where Z is a $N(0,1)$ random variable. Now,

$$\begin{aligned} P\left(\beta_j \in CI_{n,j,1-\alpha}\right) &= P\left(\left|\frac{\sqrt{n}\left(\widehat{\beta}_{n,j} - \beta_j\right)}{\sqrt{\left[\widehat{V}_n\right]_{jj}}}\right| \le z_{1-\alpha/2}\right) \\ &\to P\left(|Z| \le z_{1-\alpha/2}\right) \\ &= 1-\alpha. \end{aligned}$$

Consider, for example, the case of homoskedastic errors. We saw that in, this case, $\sqrt{n}\left(\widehat{\beta}_n-\beta\right)\rightarrow_d N\left(0,\sigma^2\left(E\left(X_iX_i'\right)\right)^{-1}\right)$. Since $s^2\rightarrow_p\sigma^2$, we can estimate the asymptotic variance by $s^2\left(n^{-1}\sum_{i=1}^n X_iX_i'\right)^{-1}$. Then, the confidence interval for β_j is given by

$$\left[\widehat{\beta}_{n,j}\pm z_{1-\alpha/2}\sqrt{\left[s^2\left(n^{-1}\sum_{i=1}^n X_iX_i'\right)^{-1}\right]_{jj}/n}\right]=\left[\widehat{\beta}_{n,j}\pm z_{1-\alpha/2}\sqrt{\left[s^2\left(X'X\right)^{-1}\right]_{jj}}\right],$$

which is the same as the finite sample confidence interval, except for the fact that we using the standard normal quantiles instead of the quantiles of the t-distribution.

Hypothesis testing

In this section we discuss asymptotic tests of the null hypothesis $H_0: h(\beta)=0$ against the alternative $H_1: h(\beta)\neq 0$, where $h: R^k\rightarrow R^q$ is a continuously differentiable function in the neighborhood of β. The restriction under H_0 includes the linear restrictions discussed in Lecture 5 as a special case (set $h(\beta)=R\beta-r$). We consider a *Wald test statistic*:

$$\begin{aligned}W_n &= nh\left(\widehat{\beta}_n\right)'\left(\widehat{AsyVar}\left(h\left(\widehat{\beta}_n\right)\right)\right)^{-1}h\left(\widehat{\beta}_n\right)\\ &= nh\left(\widehat{\beta}_n\right)'\left(\frac{\partial h\left(\widehat{\beta}_n\right)}{\partial\beta'}\widehat{V}_n\frac{\partial h\left(\widehat{\beta}_n\right)'}{\partial\beta}\right)^{-1}h\left(\widehat{\beta}_n\right),\end{aligned}$$

where $AsyVar$ denotes the asymptotic variance. The asymptotic α-size test of $H_0: h(\beta)=0$ is given by

$$\text{Reject } H_0 \text{ if } W_n>\chi^2_{q,1-\alpha},$$

where $\chi^2_{q,1-\alpha}$ is the $(1-\alpha)$ quantile of the χ^2_q distribution. A test based on W_n is called *consistent* if $P\left(W_n>\chi^2_{q,1-\alpha}|H_1\right)\rightarrow 1$.

Theorem 3 *Under Assumptions (A1), (A2*), (A6) - (A9),*

(a) $P\left(W_n>\chi^2_{q,1-\alpha}|H_0\right)\rightarrow\alpha$.

(b) $P\left(W_n>\chi^2_{q,1-\alpha}|H_1\right)\rightarrow 1$.

Proof. (a) Since $n^{1/2}\left(\widehat{\beta}_n-\beta\right)\rightarrow_d N(0,V)$ and h is continuous at β, under H_0, and by the delta method,

$$n^{1/2}h\left(\widehat{\beta}_n\right)\rightarrow_d N\left(0,\frac{\partial h(\beta)}{\partial\beta'}V\frac{\partial h(\beta)'}{\partial\beta}\right).$$

Furthermore, we have that

$$\begin{aligned}\frac{\partial h\left(\widehat{\beta}_n\right)}{\partial\beta'} &\rightarrow_p \frac{\partial h(\beta)}{\partial\beta'},\\ \widehat{V}_n &\rightarrow_p V.\end{aligned}$$

By the Cramer Convergence Theorem, under H_0,

$$\begin{aligned}\left(\frac{\partial h\left(\widehat{\beta}_n\right)}{\partial\beta'}\widehat{V}_n\frac{\partial h\left(\widehat{\beta}_n\right)'}{\partial\beta}\right)^{-1/2}n^{1/2}h\left(\widehat{\beta}_n\right) &\rightarrow_d \left(\frac{\partial h(\beta)}{\partial\beta'}V\frac{\partial h(\beta)'}{\partial\beta}\right)^{-1/2}N\left(0,\frac{\partial h(\beta)}{\partial\beta'}V\frac{\partial h(\beta)'}{\partial\beta}\right)\\ &= N\left(0,I_q\right).\end{aligned}$$

Then, by the CMT, under H_0,

$$W_n \to_d \chi^2_q,$$

which completes the proof of part (a).

(b) Under the alternative, $h(\beta) \neq 0$. Hence, by the Slutsky's Theorem,

$$\begin{aligned} h\left(\widehat{\beta}_n\right) &\to_p h(\beta) \\ &\neq 0. \end{aligned}$$

Therefore,

$$W_n/n \to_p h(\beta)' \left(\frac{\partial h(\beta)}{\partial \beta'} V \frac{\partial h(\beta)'}{\partial \beta} \right)^{-1} h(\beta),$$

and, therefore, under H_1

$$W_n \to \infty.$$

□

Note that in the case of a linear restriction $h(\beta) = R\beta - r$, we have that:

$$W_n = n\left(R\widehat{\beta}_n - r\right)' \left(R\widehat{V}_n R'\right)^{-1} \left(R\widehat{\beta}_n - r\right).$$

Further, in the homoskedastic case, one can replace $\widehat{V}_n$ by $s^2 (X'X/n)^{-1}$. Then, the Wald statistic becomes

$$W_n = \left(R\widehat{\beta}_n - r\right)' \left(s^2 R (X'X)^{-1} R'\right)^{-1} \left(R\widehat{\beta}_n - r\right),$$

which is a similar expression to that of the F statistic, except for adjustment to the number of degrees of freedom q in the numerator.

Appendix

In this section we show that $R_{1,n}$ and $R_{2,n}$ in equation (4) converge in probability to zero. The proof requires the use of the Holder's inequality.

Holder's Inequality. Let X and Y be two random variables. If $p > 1$, $q > 1$ and $1/p + 1/q = 1$, then $E|XY| \leq (E|X|^p)^{1/p} (E|Y|^q)^{1/q}$. For $p = q = 2$ one obtains the Cauchy-Schwartz inequality.

Element-by-element convergence in probability to zero is equivalent to convergence of norms to zero in probability. The norm of a matrix A is given by

$$\begin{aligned} \|A\| &= (tr(A'A))^{1/2} \\ &= \left(\sum_i \sum_j a_{ij}^2 \right)^{1/2}, \end{aligned}$$

where a_{ij} is the (i, j) element of A. For $R_{1,n}$,

$$\begin{aligned} \left\| n^{-1} \sum_{i=1}^n \left(\left(\widehat{\beta}_n - \beta\right)' X_i U_i \right) X_i X_i' \right\| &\leq n^{-1} \sum_{i=1}^n \left\| \left(\left(\widehat{\beta}_n - \beta\right)' X_i U_i \right) X_i X_i' \right\| \\ &= n^{-1} \sum_{i=1}^n tr\left(U_i^2 \left(\left(\widehat{\beta}_n - \beta\right)' X_i \right)^2 X_i X_i' X_i X_i' \right)^{1/2} \\ &= n^{-1} \sum_{i=1}^n |U_i| \left| \left(\widehat{\beta}_n - \beta\right)' X_i \right| \|X_i\| \, tr\left(X_i X_i'\right)^{1/2} \\ &= n^{-1} \sum_{i=1}^n |U_i| \left| \left(\widehat{\beta}_n - \beta\right)' X_i \right| \|X_i\|^2. \end{aligned}$$

By the Cauchy-Schwartz inequality,

$$\left|\left(\widehat{\beta}_n-\beta\right)' X_i\right| \leq \left\|\widehat{\beta}_n-\beta\right\| \|X_i\| .$$

Therefore,

$$\|R_{1,n}\| \leq \left\|\widehat{\beta}_n-\beta\right\| n^{-1}\sum_{i=1}^{n} |U_i| \, \|X_i\|^3 .$$

By the Holder's inequality with $p=4$ and $q=4/3$,

$$\begin{aligned} E\left(|U_1| \, \|X_1\|^3\right) &\leq \left(E\,|U_1|^4\right)^{1/4} \left(E\,\|X_1\|^4\right)^{3/4} \\ &< \infty, \end{aligned}$$

since by Assumption (A9) we have that $E\,|U_1|^4<\infty$, and

$$\begin{aligned} E\,\|X_1\|^4 &= E\left(\sum_{r=1}^{k} X_{1,r}^2\right)^2 \qquad (5) \\ &= \sum_{r=1}^{k}\sum_{s=1}^{k} E\left(X_{1,r}^2 X_{1,s}^2\right), \end{aligned}$$

where $E\left(X_{1,r}^2 X_{1,s}^2\right)<\infty$ due to Assumption (A8), as it was shown in the proof of Theorem 2. Hence, by the WLLN,

$$n^{-1}\sum_{i=1}^{n} |U_i| \, \|X_i\|^3 \to_p E\left(|U_1| \, \|X_1\|^3\right),$$

and since $\left\|\widehat{\beta}_n-\beta\right\| \to_p 0$, we have that $R_{1,n}\to_p 0$.

Next, consider $R_{2,n}$. By the similar argument to the one before, we can bound $R_{2,n}$ by

$$\begin{aligned} \left\| n^{-1}\sum_{i=1}^{n}\left(\left(\widehat{\beta}_n-\beta\right)' X_i\right)^2 X_i X_i' \right\| &\leq n^{-1}\sum_{i=1}^{n}\left(\left(\widehat{\beta}_n-\beta\right)' X_i\right)^2 \|X_i\| \, tr\left(X_i X_i'\right)^{1/2} \\ &= \left\|\left(\widehat{\beta}_n-\beta\right)\right\|^2 n^{-1}\sum_{i=1}^{n}\|X_i\|^4 . \end{aligned}$$

From (5) and by the WLLN,

$$n^{-1}\sum_{i=1}^{n}\|X_i\|^4 \to_p E\,\|X_1\|^4 ,$$

and, therefore, $R_{2,n}\to_p 0$.