LECTURE 8

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

This lecture discusses the large sample properties of the OLS estimator for the linear regression model defined by the following assumptions:

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

(A2*) $\mathrm{E}[U_i X_i] = 0$.

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

The large sample properties will be derived under one or more of the following additional assumptions:

(A7) $\mathrm{E}\left[X_i X_i^\top\right]$ is a finite positive definite matrix.

(A8) $\mathrm{E}[X_{i,j}^4] < \infty$ for all $j = 1, \dots, k$.

(A9) $\mathrm{E}[U_i^4] < \infty$.

(A10) $\mathrm{E}[U_i^2 X_i X_i^\top]$ 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

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

The following theorem establishes consistency of the OLS estimator.

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

Proof. The U_i's and $X_i U_i$'s are iid by Assumption (A6). Write

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

By the WLLN,[1]

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} X_i U_i &\to_p \mathrm{E}\left[X_1 U_1\right] \\ &= 0. \end{aligned}$$

Since $\mathrm{E}\left[X_1 X_1^\top\right]$ is finite, the WLLN implies that

$$n^{-1}\sum_{i=1}^{n} X_i X_i^\top \to_p \mathrm{E}\left[X_1 X_1^\top\right].$$

Since $\mathrm{E}\left[X_1 X_1^\top\right]$ is positive definite, it follows from Slutsky's Theorem that

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

Hence,

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

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

Asymptotic normality

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

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

$$\begin{aligned} V &= Q^{-1}\Omega Q^{-1}, \\ Q &= \mathrm{E}\left[X_1 X_1^\top\right], \\ \Omega &= \mathrm{E}\left[U_1^2 X_1 X_1^\top\right]. \end{aligned}$$

Proof. Rewrite (1) as

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

[1] Let X be a random variable and define

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

so that

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

Both X^+ and X^- are nonnegative random variables. When at least one of the following conditions holds: $\mathrm{E}[X^+] < \infty$ or $\mathrm{E}[X^-] < \infty$, the expected value of X is given by

$$\mathrm{E}[X] = \mathrm{E}[X^+] - \mathrm{E}[X^-].$$

The expectation $\mathrm{E}[X]$ is *not defined* when $\mathrm{E}[X^+] = \infty$ and $\mathrm{E}[X^-] = \infty$ (thus, we prohibit $\infty - \infty$). Since

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

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

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

$$\begin{aligned}
\mathrm{E}\left[\left|U_1^2 X_{1,r} X_{1,s}\right|\right] &\leq \left(\mathrm{E}[U_1^4]\,\mathrm{E}\left[X_{1,r}^2 X_{1,s}^2\right]\right)^{1/2} \\
&\leq \left(\mathrm{E}[U_1^4]\right)^{1/2}\left(\mathrm{E}[X_{1,r}^4]\,\mathrm{E}[X_{1,s}^4]\right)^{1/4} \\
&< \infty.
\end{aligned}$$

By the CLT,

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

Finally, it follows from (2), (3) and the Cramér Convergence Theorem (multivariate version), that

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

□

Remarks:

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

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

$$V = \left(\mathrm{E}\left[X_1 X_1^\top\right]\right)^{-1} \mathrm{E}\left[U_1^2 X_1 X_1^\top\right] \left(\mathrm{E}\left[X_1 X_1^\top\right]\right)^{-1}.$$

If Assumption (A3) holds, so that $\mathrm{E}\left[U_1^2 \mid X_1\right] = \sigma^2$, then V simplifies to $\sigma^2\left(\mathrm{E}\left[X_1 X_1^\top\right]\right)^{-1}$. By the LIE,

$$\begin{aligned}
\mathrm{E}\left[U_1^2 X_1 X_1^\top\right] &= \mathrm{E}\left[\mathrm{E}\left[U_1^2 X_1 X_1^\top \mid X_1\right]\right] \\
&= \mathrm{E}\left[X_1 X_1^\top\,\mathrm{E}\left[U_1^2 \mid X_1\right]\right] \\
&= \sigma^2\,\mathrm{E}\left[X_1 X_1^\top\right].
\end{aligned}$$

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 $\widehat{U}_i = Y_i - X_i^\top \widehat{\beta}_n$. Consider the following estimator of V implied by the MM principle:

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

As shown above, $\widehat{Q}_n^{-1} \to_p Q^{-1}$. Next, consider $\widehat{\Omega}_n$ and note that $\widehat{U}_i$ depends on all U_i's, $i = 1, \ldots, n$, through $\widehat{\beta}_n$, so the WLLN cannot be applied directly to averages involving $\widehat{U}_i$. For $\widehat{\Omega}_n$, one cannot directly apply properties (ii) and (iii) of probability limits from Lecture 7, because the number of summands grows with n. Write

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

Therefore,

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

Under the assumptions of Theorem 2, $\mathrm{E}\left[U_i^2 X_i X_i^\top\right]$ is finite, as was shown in its proof. Consequently, by the WLLN,

$$n^{-1} \sum_{i=1}^n U_i^2 X_i X_i^\top \to_p \mathrm{E}\left[U_1^2 X_1 X_1^\top\right].$$

Both $R_{1,n}$ 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 estimator in the cases of homoskedastic or heteroskedastic errors. It is often called the robust, heteroskedasticity-consistent, or White estimator (White, 1980, *Econometrica*). Many statistical software packages can compute standard errors using the White estimator; by default, however, they report standard errors based on the homoskedastic estimator $s^2 \left(n^{-1} \sum_{i=1}^n X_i X_i^\top \right)^{-1}$ (here s^2 is a sequence indexed by n, and one can show that $s^2 \to_p \sigma^2 = \mathrm{E}[U_1^2]$).

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 $\Pr\left(\beta_j \in CI_{n,j,1-\alpha}\right) \to 1-\alpha$ 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 Slutsky's Theorem and the Cramér Convergence Theorem 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 also be written as

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

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

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

Consider, for example, the case of homoskedastic errors. We saw that in this case, $\sqrt{n}\left(\widehat{\beta}_n - \beta\right) \to_d N\left(0, \sigma^2\left(\mathrm{E}\left[X_1 X_1^\top\right]\right)^{-1}\right)$. Since $s^2 \to_p \sigma^2$, we can estimate the asymptotic variance by $s^2\left(n^{-1}\sum_{i=1}^n X_i X_i^\top\right)^{-1}$. Then, the confidence interval for β_j is given by

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

which is the same as the finite sample confidence interval, except for the fact that we are 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 $h : \mathbb{R}^k \to \mathbb{R}^q$ is continuously differentiable in a 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*:

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

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 $\Pr\left(W_n > \chi^2_{q,1-\alpha} \mid H_1\right) \to 1$.

Theorem 3. *Under Assumptions (A1), (A2*), (A6)–(A10),*

(a) $\Pr\left(W_n > \chi^2_{q,1-\alpha} \mid H_0\right) \to \alpha$*.*

(b) $\Pr\left(W_n > \chi^2_{q,1-\alpha} \mid H_1\right) \to 1$*.*

Proof. (a) Under H_0, since $n^{1/2}\left(\widehat{\beta}_n - \beta\right) \to_d N(0, V)$ and h is continuously differentiable at β, the delta method gives

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

Furthermore,

$$\frac{\partial h\left(\widehat{\beta}_n\right)}{\partial \beta^\top} \to_p \frac{\partial h(\beta)}{\partial \beta^\top},$$
$$\widehat{V}_n \to_p V.$$

By the Cramér Convergence Theorem, under H_0,

$$\left(\frac{\partial h\left(\widehat{\beta}_n\right)}{\partial \beta^\top} \widehat{V}_n \frac{\partial h\left(\widehat{\beta}_n\right)^\top}{\partial \beta}\right)^{-1/2} n^{1/2} h\left(\widehat{\beta}_n\right) \to_d \left(\frac{\partial h(\beta)}{\partial \beta^\top} V \frac{\partial h(\beta)^\top}{\partial \beta}\right)^{-1/2} N\left(0, \frac{\partial h(\beta)}{\partial \beta^\top} V \frac{\partial h(\beta)^\top}{\partial \beta}\right)$$
$$= N(0, I_q).$$

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 Slutsky's Theorem,

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

Therefore,

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

and, therefore, under H_1,

$$W_n \to \infty.$$

□

In the case of a linear restriction $h(\beta) = R\beta - r$,

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

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

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

which resembles the F-statistic from Lecture 5 but without the division by q (the number of restrictions).

Appendix

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

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

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

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

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

By the Cauchy-Schwarz inequality,

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

Therefore,

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

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

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

since $\mathrm{E}\left[|U_1|^4\right] < \infty$ by Assumption (A9), and

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

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

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

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

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

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

From (5) and by the WLLN,

$$n^{-1} \sum_{i=1}^{n} \left\|X_i\right\|^4 \to_p \mathrm{E}\left[\left\|X_1\right\|^4\right],$$

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