LECTURE 4
CONFIDENCE INTERVALS

In this lecture, we consider the normal regression model defined by Assumptions (A1)–(A5).

The *point estimator* $\widehat{\beta}$ of the vector of parameters β is not very informative, since $\Pr(\widehat{\beta} = \beta) = 0$. We construct *random* intervals or regions that contain the true parameter with a specified probability $1-\alpha$, where α is a small number ($1-\alpha$ is called the *confidence level*). Common choices are $\alpha = 0.01$, 0.05, and 0.10. A confidence interval with *coverage probability* $1-\alpha$ is denoted $CI_{1-\alpha}$.

Scalar case

Suppose that we are interested in constructing a confidence interval for β_1, the first element of β. The partitioned regression is given by

$$Y = \beta_1 X_1 + X_2\beta_2 + U,$$

where X_1 is the $n \times 1$ vector of observations on the first regressor. The OLS estimator of β_1 is

$$\widehat{\beta}_1 = \frac{X_1^\top M_2 Y}{X_1^\top M_2 X_1},$$

where $M_2 = I_n - X_2\left(X_2^\top X_2\right)^{-1} X_2^\top$.

A natural approach is to consider *symmetric* intervals centered at the point estimator:

$$CI_{1-\alpha} = \left[\widehat{\beta}_1 - c, \widehat{\beta}_1 + c\right]. \tag{1}$$

Since $\widehat{\beta}_1$ is a function of the random sample, the confidence interval $CI_{1-\alpha}$ as defined in (1) is random as well. We choose c so that

$$\Pr\left(\beta_1 \in CI_{1-\alpha} \mid X\right) = 1-\alpha,$$

where $X = (X_1 \quad X_2)$. Finding c requires the conditional distribution of $\widehat{\beta}_1$ given X. Under Assumptions (A1)–(A5),

$$\widehat{\beta}_1 \mid X \sim N\left(\beta_1, \sigma^2/\left(X_1^\top M_2 X_1\right)\right), \text{ and, consequently,}$$

$$\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/\left(X_1^\top M_2 X_1\right)}} \mid X \sim N(0,1). \tag{2}$$

To verify this, recall that $\widehat{\beta}_1$ is a linear estimator and write $\widehat{\beta}_1 = \beta_1 + \left(X_1^\top M_2 U\right)/\left(X_1^\top M_2 X_1\right)$.

Let z_τ denote the τ-quantile of the standard normal distribution: if $Z \sim N(0,1)$, then

$$\Pr\left(Z \le z_\tau\right) = \tau.$$

For example, for $\tau = 0.5$ we have the *median*:

$$\Pr\left(Z \le z_{0.5}\right) = 0.5.$$

Since the standard normal distribution is symmetric about zero,

$$z_\alpha = -z_{1-\alpha}, \text{ and, therefore,}$$

$$\Pr\left(-z_{1-\alpha/2} \le Z \le z_{1-\alpha/2}\right) = 1-\alpha.$$

For example, for $\alpha = 0.05$, $z_{1-0.05/2} = z_{0.975} = 1.96$, and $z_{0.025} = -1.96$.

σ^2 is known

Suppose σ^2 is known, so the variance of $\widehat{\beta}_1$ can be computed exactly. We set

$$\begin{aligned} c &= z_{1-\alpha/2}\sqrt{\operatorname{Var}(\widehat{\beta}_1 \mid X)} \\ &= z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1^\top M_2 X_1\right)}. \end{aligned}$$

We show that

$$\Pr\left(\beta_1 \in \left[\widehat{\beta}_1 - z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1^\top M_2 X_1\right)}, \widehat{\beta}_1 + z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1^\top M_2 X_1\right)}\right] \mid X\right) = 1-\alpha.$$

To see this, write

$$\begin{aligned} &\Pr\left(\widehat{\beta}_1 - z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1^\top M_2 X_1\right)} \leq \beta_1 \leq \widehat{\beta}_1 + z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1^\top M_2 X_1\right)} \mid X\right) \\ &= \Pr\left(-z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1^\top M_2 X_1\right)} \leq \beta_1 - \widehat{\beta}_1 \leq z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1^\top M_2 X_1\right)} \mid X\right) \\ &= \Pr\left(-z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1^\top M_2 X_1\right)} \leq \widehat{\beta}_1 - \beta_1 \leq z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1^\top M_2 X_1\right)} \mid X\right) \\ &= \Pr\left(-z_{1-\alpha/2} \leq \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/\left(X_1^\top M_2 X_1\right)}} \leq z_{1-\alpha/2} \mid X\right). \end{aligned} \tag{3}$$

The desired result follows immediately from (2), (3), and the definition of $z_{1-\alpha/2}$.

σ^2 is unknown

The construction above requires σ^2 to be known. When σ^2 is unknown, we replace it with an estimator:

$$s^2 = \widehat{U}^\top \widehat{U}/(n-k).$$

However, $(\widehat{\beta}_1 - \beta_1)/\sqrt{s^2/(X_1^\top M_2 X_1)}$ is not normally distributed, since the denominator is random. Consequently, standard normal quantiles cannot be used to construct confidence intervals.

The key result is that

$$\frac{\left(\widehat{\beta}_1 - \beta_1\right)}{\sqrt{s^2/\left(X_1^\top M_2 X_1\right)}} \mid X \sim t_{n-k}. \tag{4}$$

The t_{n-k} distribution is the distribution of

$$Z/\sqrt{V/(n-k)},$$

where Z is a standard normal random variable, V is a χ^2_{n-k} random variable, and Z and V are independent. We write

$$\begin{aligned} \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{s^2/\left(X_1^\top M_2 X_1\right)}} &= \left(\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/\left(X_1^\top M_2 X_1\right)}}\right)/\sqrt{\frac{s^2}{\sigma^2}} \\ &= \left(\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/\left(X_1^\top M_2 X_1\right)}}\right)/\sqrt{\frac{\widehat{U}^\top \widehat{U}}{\sigma^2}/(n-k)}. \end{aligned} \tag{5}$$

We already know that, in the above expression, $\left(\widehat{\beta}_1 - \beta_1\right) / \sqrt{\sigma^2 / \left(X_1^\top M_2 X_1\right)} \mid X \sim N(0,1)$. We show next that, conditional on X,

$$\frac{\widehat{U}^\top \widehat{U}}{\sigma^2} \mid X \sim \chi^2_{n-k}. \tag{6}$$

To show this, we need the following result:

Lemma 1. *Suppose the n-vector* $U \sim N(0, I_n)$. *Let* A *be a non-random* $n \times n$ *symmetric and idempotent matrix with* $\text{rank}(A) = r$. *Then,* $U^\top A U \sim \chi^2_r$.

Proof. It suffices to show that $U^\top A U = \sum_{i=1}^r Z_i^2$, where Z_i are iid $N(0,1)$.

Since A is a symmetric matrix, one can write

$$A = C \Lambda C^\top,$$

where Λ is an $n \times n$ diagonal matrix consisting of the eigenvalues of A, and $C^\top C = I_n$. Since A is idempotent,

$$\begin{aligned} A &= AA, \text{ and} \\ C\Lambda C^\top &= \left(C\Lambda C^\top\right)\left(C\Lambda C^\top\right) \\ &= C\Lambda^2 C^\top. \end{aligned}$$

Therefore,

$$\Lambda = \Lambda^2,$$

which implies that each eigenvalue of A is either zero or one. Since the rank of a matrix equals the number of its non-zero eigenvalues, there are r non-zero eigenvalues λ_i in

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

Define

$$Z = C^\top U.$$

Since $U \sim N(0, I_n)$,

$$\begin{aligned} Z &\sim N\left(0, C^\top C\right) \\ &\sim N(0, I_n). \end{aligned}$$

Lastly,

$$\begin{aligned} U^\top A U &= Z^\top \Lambda Z \\ &= \sum_{i=1}^n \lambda_i Z_i^2. \end{aligned}$$

The result follows because Z_i are iid $N(0,1)$, there are r eigenvalues equal to one, and the remaining $n-r$ eigenvalues are zero. □

Now, to show (6), write

$$\frac{\widehat{U}^\top \widehat{U}}{\sigma^2} = \left(\frac{U}{\sigma}\right)^\top M_X \left(\frac{U}{\sigma}\right), \tag{7}$$

where

$$M_X = I_n - X\left(X^\top X\right)^{-1} X^\top.$$

By Assumption (A5),

$$\frac{U}{\sigma} \mid X \sim N(0, I_n). \tag{8}$$

Since M_X is symmetric and idempotent, its eigenvalues are either zero or one. Therefore,

$$\begin{aligned}\operatorname{rank}(M_X) &= \operatorname{tr}(M_X) \\ &= n - k. \end{aligned} \tag{9}$$

The result in (6) follows from (7), (8), (9), and Lemma 1.

Finally, we show that $\widehat{\beta}_1 - \beta_1$ and $\widehat{U}^\top \widehat{U}$ in (5) are independent given X. Write

$$\begin{aligned}\widehat{\beta}_1 - \beta_1 &= \left(X_1^\top M_2 U\right) / \left(X_1^\top M_2 X_1\right), \\ \widehat{U}^\top \widehat{U} &= U^\top M_X U.\end{aligned}$$

It is sufficient to show independence of $X_1^\top M_2 U$ and $M_X U$. Since $\widehat{\beta}_1$ is a function of $X_1^\top M_2 U$, and $\widehat{U}^\top \widehat{U}$ is a function of $M_X U$, independence of $X_1^\top M_2 U$ and $M_X U$ implies independence of $\widehat{\beta}_1$ and $\widehat{U}^\top \widehat{U}$. First, we show that they are uncorrelated:

$$\begin{aligned}\operatorname{Cov}(X_1^\top M_2 U, M_X U \mid X) &= \mathrm{E}\left[X_1^\top M_2 U U^\top M_X \mid X\right] \\ &= X_1^\top M_2 \,\mathrm{E}\left[U U^\top \mid X\right] M_X \\ &= X_1^\top M_2 \left(\sigma^2 I_n\right) M_X \\ &= \sigma^2 X_1^\top M_2 M_X \\ &= \sigma^2 X_1^\top M_X \text{ (since } M_2 M_X = M_X\text{, see Lecture 3)} \\ &= 0.\end{aligned}$$

Since $X_1^\top M_2 U$ and $M_X U$ are both linear functions of U, they are jointly normal conditional on X. For jointly normal random vectors, zero correlation implies independence. Consequently, $\widehat{\beta}_1 - \beta_1$, a function of $X_1^\top M_2 U$, and $\widehat{U}^\top \widehat{U}$ are independent as well.

This establishes (4). Consequently, when constructing confidence intervals, if one replaces the unknown σ^2 with s^2, one must replace $z_{1-\alpha/2}$ with quantiles of the t distribution, $t_{n-k,1-\alpha/2}$:

$$CI_{1-\alpha} = \left[\widehat{\beta}_1 - t_{n-k,1-\alpha/2}\sqrt{s^2/\left(X_1^\top M_2 X_1\right)}, \widehat{\beta}_1 + t_{n-k,1-\alpha/2}\sqrt{s^2/\left(X_1^\top M_2 X_1\right)}\right].$$

The expression $s^2/\left(X_1^\top M_2 X_1\right)$ that appears in the above equation is the *estimated* variance of $\widehat{\beta}_1$:

$$\widehat{\operatorname{Var}}(\widehat{\beta}_1 \mid X) = s^2/\left(X_1^\top M_2 X_1\right).$$

A $(1-\alpha)$-level confidence interval for β_j, $j = 1, \ldots, k$, is:

$$CI_{1-\alpha}^j = \left[\widehat{\beta}_j - t_{n-k,1-\alpha/2}\sqrt{\widehat{\operatorname{Var}}(\widehat{\beta}_j \mid X)}, \widehat{\beta}_j + t_{n-k,1-\alpha/2}\sqrt{\widehat{\operatorname{Var}}(\widehat{\beta}_j \mid X)}\right]. \tag{10}$$

Vector case

Consider the full parameter vector $\beta = (\beta_1, \beta_2, \ldots, \beta_k)^\top$. Equation (10) describes how to construct individual confidence intervals for the elements of β. These are concerned with *marginal* distributions of the elements of β, and their Cartesian product does not contain β with the desired probability. In this section, we construct random *regions* that contain β with a specified probability $1-\alpha$. We retain the notation $CI_{1-\alpha}$, although now $CI_{1-\alpha} \subseteq \mathbb{R}^k$.

The following is a simple and conventional approach to constructing confidence regions. We are looking for a *random* region $CI_{1-\alpha} \subseteq \mathbb{R}^k$ such that $\Pr\left(\beta \in CI_{1-\alpha} \mid X\right) = 1 - \alpha$. Consider a quadratic form in $\left(\widehat{\beta} - \beta\right)$:

$$\begin{aligned}
&\left(\widehat{\beta} - \beta\right)^\top \left(\widehat{\mathrm{Var}}(\widehat{\beta} \mid X)\right)^{-1} \left(\widehat{\beta} - \beta\right)/k \\
&= \left(\widehat{\beta} - \beta\right)^\top \left(s^2 \left(X^\top X\right)^{-1}\right)^{-1} \left(\widehat{\beta} - \beta\right)/k \\
&= \frac{\left(\widehat{\beta} - \beta\right)^\top \left(\sigma^2 \left(X^\top X\right)^{-1}\right)^{-1} \left(\widehat{\beta} - \beta\right)/k}{s^2/\sigma^2} \\
&= \frac{\left(\widehat{\beta} - \beta\right)^\top \left(\sigma^2 \left(X^\top X\right)^{-1}\right)^{-1} \left(\widehat{\beta} - \beta\right)/k}{\left(\frac{\widehat{U}^\top \widehat{U}}{\sigma^2}\right)/(n-k)}.
\end{aligned} \tag{11}$$

Next, we show that the expression in (11) has the $F_{k,n-k}$ distribution given X.

The $F_{k,n-k}$ distribution is defined as the distribution of

$$\frac{V/k}{W/(n-k)},$$

where $V \sim \chi^2_k$, $W \sim \chi^2_{n-k}$, and V and W are independent. From the discussion in the previous section, we know that $\widehat{U}^\top \widehat{U}/\sigma^2 \mid X \sim \chi^2_{n-k}$ and that it is independent of the numerator in (11). Thus, we need to show that

$$\left(\widehat{\beta} - \beta\right)^\top \left(\sigma^2 \left(X^\top X\right)^{-1}\right)^{-1} \left(\widehat{\beta} - \beta\right) \mid X \sim \chi^2_k. \tag{12}$$

We need the following lemma.

Lemma 2. *Suppose the k-vector $U \sim N(0, \Sigma)$, where Σ is a positive definite variance-covariance matrix. Then, $U^\top \Sigma^{-1} U \sim \chi^2_k$.*

Proof. Since Σ is symmetric, $\Sigma = C\Lambda C^\top$, where Λ is a diagonal matrix of the eigenvalues of Σ on its main diagonal, and $C^\top C = CC^\top = I_k$. Since Σ is positive definite, its eigenvalues are positive, and therefore, $\Lambda^{1/2}$ can be defined as

$$\Lambda^{1/2} = \begin{pmatrix} \lambda_1^{1/2} & & 0 \\ & \ddots & \\ 0 & & \lambda_k^{1/2} \end{pmatrix},$$

and $\Lambda^{-1/2}$ can be defined as

$$\Lambda^{-1/2} = \begin{pmatrix} \lambda_1^{-1/2} & & 0 \\ & \ddots & \\ 0 & & \lambda_k^{-1/2} \end{pmatrix}.$$

Next, since $C\Lambda^{-1}C^\top C\Lambda C^\top = I_k$, we have that

$$\Sigma^{-1} = C\Lambda^{-1}C^\top.$$

Now, define

$$\Sigma^{1/2} = C\Lambda^{1/2}C^\top \text{ and } \Sigma^{-1/2} = C\Lambda^{-1/2}C^\top.$$

Both $\Sigma^{1/2}$ and $\Sigma^{-1/2}$ are symmetric: $\left(\Sigma^{1/2}\right)^\top = \Sigma^{1/2}$ and $\left(\Sigma^{-1/2}\right)^\top = \Sigma^{-1/2}$. Furthermore,

$$\begin{aligned}
\Sigma^{1/2}\Sigma^{1/2} &= C\Lambda^{1/2}C^\top C\Lambda^{1/2}C^\top = C\Lambda^{1/2}\Lambda^{1/2}C^\top = C\Lambda C^\top = \Sigma, \\
\Sigma^{-1/2}\Sigma\Sigma^{-1/2} &= C\Lambda^{-1/2}C^\top C\Lambda C^\top C\Lambda^{-1/2}C^\top = C\Lambda^{-1/2}\Lambda\Lambda^{-1/2}C^\top = CC^\top = I_k.
\end{aligned}$$

The matrix $\Sigma^{1/2}$ is called the symmetric square root of Σ, and $\Sigma^{-1/2}$ is the inverse symmetric square root. Define a k-vector

$$V = \Sigma^{-1/2} U,$$

so that

$$U^\top \Sigma^{-1} U = V^\top V. \tag{13}$$

Since $U \sim N(0, \Sigma)$ and V is a linear transformation of U,

$$\begin{aligned} V &\sim N\left(0, \Sigma^{-1/2} \operatorname{Var}(U) \Sigma^{-1/2}\right) \\ &= N\left(0, \Sigma^{-1/2} \Sigma \Sigma^{-1/2}\right) \\ &= N(0, I_k). \end{aligned}$$

From (13) and the definition of the χ^2_k distribution,

$$U^\top \Sigma^{-1} U = V^\top V = \sum_{j=1}^{k} V_j^2 \sim \chi^2_k.$$

□

Now, the result in (12) follows from Lemma 2. Consequently,

$$\frac{\left(\widehat{\beta} - \beta\right)^\top \left(s^2 \left(X^\top X\right)^{-1}\right)^{-1} \left(\widehat{\beta} - \beta\right)}{k} \mid X \sim F_{k,n-k}.$$

Let $F_{k,n-k,\tau}$ be the τ-quantile of the $F_{k,n-k}$ distribution. We construct the $(1-\alpha)$-level confidence region as follows:

$$CI_{1-\alpha} = \left\{ b \in \mathbb{R}^k : \left(\widehat{\beta} - b\right)^\top \left(s^2 \left(X^\top X\right)^{-1}\right)^{-1} \left(\widehat{\beta} - b\right) / k \leq F_{k,n-k,1-\alpha} \right\}.$$

From the above discussion, it follows that

$$\begin{aligned} &\Pr\left(\beta \in CI_{1-\alpha} \mid X\right) \\ &= \Pr\left(\left(\widehat{\beta} - \beta\right)^\top \left(s^2 \left(X^\top X\right)^{-1}\right)^{-1} \left(\widehat{\beta} - \beta\right) / k \leq F_{k,n-k,1-\alpha} \mid X \right) \\ &= 1 - \alpha. \end{aligned}$$

Remark

The confidence interval/region $CI_{1-\alpha}$ is a function of the sample $\{(Y_i, X_i) : i = 1, \ldots, n\}$, and is therefore random, which allows us to talk about the probability of $CI_{1-\alpha}$ containing the true value of β. On the other hand, the realization of $CI_{1-\alpha}$ is not random. Once the confidence interval is computed from the data, it is no longer meaningful to speak of the probability that it contains β: the true parameter either lies in the interval or it does not.