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 $P\left(\widehat{\beta}=\beta\right)=0$. In this lecture, we consider construction of *random* intervals or regions with the property that they include the true parameter with some specified probability $1-\alpha$, where α is a small number ($1-\alpha$ is called *confidence level).* Usually, the following values of α are considered: 0.01, 0.05, 0.10. A confidence interval with *coverage probability* $1-\alpha$ is denoted by CI_{1-a}.

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, in this case, X_1 is a $n\times 1$ vector that contains all n observations for the first regressor. The OLS estimator of β_1 is

$$\widehat{\beta}_1=\frac{X_1'M_2Y}{X_1'M_2X_1},$$

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

One method of constructing confidence intervals is to consider *symmetric* intervals around the point estimator:

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

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. The problem is to choose c so that

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

where $X=\begin{pmatrix} X_1 & X_2\end{pmatrix}$. In order to find c, one has to know the conditional distribution of $\widehat{\beta}_1$ given X. Under Assumptions (A1)-(A5),

$$\widehat{\beta}_1|X\sim N\left(\beta_1,\sigma^2/\left(X_1'M_2X_1\right)\right), \text{ and, consequently,}$$
$$\frac{\widehat{\beta}_1-\beta_1}{\sqrt{\sigma^2/\left(X_1'M_2X_1\right)}}|X\sim N\left(0,1\right). \quad (2)$$

In order to show that, note that $\widehat{\beta}_1$ is a linear estimator, and write $\widehat{\beta}_1=\beta_1+\left(X_1'M_2U\right)/\left(X_1'M_2X_1\right)$.

Let z_τ be the τ-*quantile* of the standard normal distribution; that is, if $Z\sim N(0,1)$, then

$$P\left(Z\leq z_\tau\right)=\tau.$$

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

$$P\left(Z\leq z_{0.5}\right)=0.5.$$

Note that, since the standard normal distribution is symmetric around zero, we have that

$$z_\alpha=-z_{1-\alpha}, \text{ and, therefore,}$$
$$P\left(-z_{1-\alpha/2}\leq Z\leq z_{1-\alpha/2}\right)=1-\alpha.$$

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 for a moment that σ^2 is known and that we can actually compute the variance of $\widehat{\beta}_1$. We set

$$\begin{aligned} c &= z_{1-\alpha/2}\sqrt{Var\left(\widehat{\beta}_1|X\right)} \\ &= z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1'M_2X_1\right)}. \end{aligned}$$

We will show that

$$P\left(\beta_1 \in \left[\widehat{\beta}_1 - z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1'M_2X_1\right)}, \widehat{\beta}_1 + z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1'M_2X_1\right)}\right]|X\right) = 1-\alpha.$$

Indeed,

$$\begin{aligned} &P\left(\widehat{\beta}_1 - z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1'M_2X_1\right)} \le \beta_1 \le \widehat{\beta}_1 + z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1'M_2X_1\right)}|X\right) \\ &= P\left(-z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1'M_2X_1\right)} \le \beta_1 - \widehat{\beta}_1 \le z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1'M_2X_1\right)}|X\right) \\ &= P\left(-z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1'M_2X_1\right)} \le \widehat{\beta}_1 - \beta_1 \le z_{1-\alpha/2}\sqrt{\sigma^2/\left(X_1'M_2X_1\right)}|X\right) \\ &= P\left(-z_{1-\alpha/2} \le \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/\left(X_1'M_2X_1\right)}} \le z_{1-\alpha/2}|X\right). \qquad (3) \end{aligned}$$

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

σ^2 is unknown

Construction of the $CI_{1-\alpha}$ above relies on the fact that σ^2 is known. In the case of the unknown σ^2, one can take a similar approach, first replacing σ^2 with its estimator, for example:

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

However, $\left(\widehat{\beta}_1 - \beta_1\right)/\sqrt{s^2/\left(X_1'M_2X_1\right)}$ is not normally distributed, since it is a nonlinear function of the random $\widehat{\beta}_1$ and s^2. Therefore, one cannot use the quantiles of the standard normal distribution for construction of confidence intervals.

It turns out that

$$\frac{\left(\widehat{\beta}_1 - \beta_1\right)}{\sqrt{s^2/\left(X_1'M_2X_1\right)}}|X \sim t_{n-k}. \qquad (4)$$

Recall that t_{n-k} distribution is defined as follows:

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

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

$$\begin{aligned} \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{s^2/\left(X_1'M_2X_1\right)}} &= \left(\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/\left(X_1'M_2X_1\right)}}\right)/\sqrt{\frac{s^2}{\sigma^2}} \\ &= \left(\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/\left(X_1'M_2X_1\right)}}\right)/\sqrt{\frac{\widehat{U}'\widehat{U}}{\sigma^2}/(n-k)}. \qquad (5) \end{aligned}$$

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

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

In order to show that we need the following result:

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

Proof. It is sufficient to show that $U'AU = \sum_i^r Z_i^2$, where Z_i are iid $N(0,1)$.

Since A is symmetric matrix, one can write

$$A = C\Lambda C',$$

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

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

Therefore,

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

which implies that all eigenvalues of A are either zeros or ones. Since the rank of a matrix equals the number of its non-zero eigenvalues, there are r non-zero eigenvalues λ_i in

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

Define

$$Z = C'U.$$

Since $U \sim N(0, I_n)$, we have that

$$\begin{aligned} Z &\sim N(0, C'C) \\ &\sim N(0, I_n). \end{aligned}$$

Lastly,

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

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

Now, to show (6), write

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

where

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

By Assumption (A5),

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

Since M_X is symmetric and idempotent, its eigenvalues are either zeros or ones. Therefore,

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

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

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

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

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

$$\begin{aligned} Cov\,(X_1' M_2 U, M_X U|X) &= E\,(X_1' M_2 U U' M_X|X) \\ &= X_1' M_2 E\,(UU'|X)\, M_X \\ &= X_1' M_2 \left(\sigma^2 I_n\right) M_X \\ &= \sigma^2 X_1' M_2 M_X \\ &= \sigma^2 X_1' M_X \text{ (see Lecture 3, page 8)} \\ &= 0. \end{aligned}$$

Since $X_1' M_2 U$ and $M_X U$ are linear functions of U, they are normal conditional on X. Since they are uncorrelated, normality implies that they are independent. Consequently, $\widehat{\beta}_1 - \beta_1$, a function of $X_1' M_2 U$, and $\widehat{U}'\widehat{U}$ are independent as well.

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

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

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

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

Thus, one constructs a level α confidence interval for β_j, $j = 1, \ldots, k$ as follows

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

Vector case

Suppose that we are concerned with the vector of parameters $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$. 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 simple combination does not produce a set that includes the whole vector β with a desired probability. In this section, we consider construction of random *regions* that include β with some specified probability $1-\alpha$. We keep the notation $CI_{1-\alpha}$, despite the fact that in this case $CI_{1-\alpha}$ is now subset of R^k.

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

$$\begin{aligned}
&\left(\widehat{\beta}-\beta\right)'\left(\widehat{Var}(\hat{\beta}|X)\right)^{-1}\left(\widehat{\beta}-\beta\right)/k \\
&= \left(\widehat{\beta}-\beta\right)'\left(s^2\left(X'X\right)^{-1}\right)^{-1}\left(\widehat{\beta}-\beta\right)/k \\
&= \frac{\left(\widehat{\beta}-\beta\right)'\left(\sigma^2\left(X'X\right)^{-1}\right)^{-1}\left(\widehat{\beta}-\beta\right)/k}{s^2/\sigma^2} \\
&= \frac{\left(\widehat{\beta}-\beta\right)'\left(\sigma^2\left(X'X\right)^{-1}\right)^{-1}\left(\widehat{\beta}-\beta\right)/k}{\left(\frac{\widehat{U}'\widehat{U}}{\sigma^2}\right)/(n-k)}. \qquad (11)
\end{aligned}$$

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 independent. From the discussion in the previous section we know that $\widehat{U}'\widehat{U}/\sigma^2|X \sim \chi^2_{n-k}$ and independent from the numerator in (11). Thus, we need to show that

$$\left(\widehat{\beta}-\beta\right)'\left(\sigma^2\left(X'X\right)^{-1}\right)^{-1}\left(\widehat{\beta}-\beta\right)|X \sim \chi^2_k. \qquad (12)$$

We need the following lemma.

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

Proof. Since Σ is symmetric, $\Sigma = C\Lambda C'$, where Λ is a diagonal matrix of the eigenvalues of Σ on its main diagonal, and $C'C = CC' = I_k$. 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 $C\Lambda^{-1}C'C\Lambda C' = I_k$, we have that

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

Now, define

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

We have that $\left(\Sigma^{1/2}\right)' = \Sigma^{1/2}$ and $\left(\Sigma^{-1/2}\right)' = \Sigma^{-1/2}$ (symmetric). Furthermore,

$$\Sigma^{1/2}\Sigma^{1/2} = C\Lambda^{1/2}C'C\Lambda^{1/2}C' = C\Lambda^{1/2}\Lambda^{1/2}C' = C\Lambda C' = \Sigma,$$
$$\Sigma^{-1/2}\Sigma\Sigma^{-1/2} = C\Lambda^{-1/2}C'C\Lambda C'C\Lambda^{-1/2}C' = C\Lambda^{-1/2}\Lambda\Lambda^{-1/2}C' = CC' = I_k.$$

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

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

so that

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

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

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

Thus, from (13) and by the definition of the χ^2_k,

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

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

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

Let $F_{k,n-k,\tau}$ be the τ-quantile of the F distribution. We construct the α-level confidence region as follows

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

From the above discussion it follows that

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

Remark

The confidence interval/region $CI_{1-\alpha}$ is a function of the sample $\{(Y_i, X_i) : i = 1, \ldots, n\}$, and therefore random, which allows us to talk about 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 given the data, it does not make sense anymore to talk about the probability that it includes β. It is either zero or one.