Economics 522A, Spring 2007

Lecture Note 9: Inference for β and σ^2 under Normality

Distribution of s^2:

To get some intuition, suppose for a moment that β is known. Then we could form residuals without using the OLS estimator $\hat{\beta}$:

$$\epsilon = y - X\beta = \begin{pmatrix} y_1 - x_1'\beta \\ \vdots \\ y_n - x_2'\beta \end{pmatrix} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Then

$$\epsilon \mid X \sim N(0, \sigma^2 I_n).$$
$$\Rightarrow \frac{\epsilon}{\sigma} \mid X \sim N(0, I_n).$$

So

$$\frac{\epsilon'\epsilon}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{\epsilon_i}{\sigma}\right)^2$$

is the sum of n squared, independent, standard normal random variables. Therefore

$$\frac{\epsilon'\epsilon}{\sigma^2} \mid X \sim \chi^2_n,$$

and, since the distribution on the right does not depend on X, the term $\epsilon'\epsilon/\sigma^2$ is marginally distributed as χ^2_n.

Now, if we do not know β, we cannot directly form ϵ. Instead we first calculate $\hat{\beta}$ and then form $e = y - X\hat{\beta}$. This does have an effect on the distribution of e, but it turns out that a similar result holds:

$$\frac{e'e}{\sigma^2} \sim \chi^2_{n-k}.$$

(For a proof of this result, see Ruud.)

Since $s^2 = (e'e)/(n-k)$,

$$\frac{(n-k)s^2}{\sigma^2} \sim \chi^2_{n-k}.$$

Alternative, we can write this as

$$s^2 \sim \chi^2_{n-k} \cdot \left(\frac{\sigma^2}{n-k}\right).$$

Recall that in LN8, we also showed that s^2 is independent of $\hat{\beta}$.

Inference[1] for σ^2:

We would like to form confidence intervals for σ^2, and test hypotheses about σ^2. Because s^2 has a distribution that depends on σ^2 but not on β, we can do so without knowing β. (We say that s^2 is "pivotal," because its distribution does not depend on the other unknown parameters besides σ^2.)

Confidence Interval:

For a given $\alpha \in (0,1)$, for example $\alpha = 0.05$, let c_0, c_1 satisfy:

$$P\left(c_0 \le \chi^2_{n-k} \le c_1\right) = 1 - \alpha.$$

Then

$$\begin{aligned} 1-\alpha &= P\left(c_0 \le \frac{(n-k)s^2}{\sigma^2} \le c_1\right) \\ &= P\left(\frac{(n-k)s^2}{c_1} \le \sigma^2 \le \frac{(n-k)s^2}{c_0}\right). \end{aligned}$$

So the interval

$$\left[\frac{(n-k)s^2}{c_1}, \frac{(n-k)s^2}{c_0}\right]$$

will be a $(1-\alpha)$ confidence interval for σ^2.

Note: for a given α, the numbers c_0, c_1 are not uniquely determined: there are multiple possible pairs of values satisfying $P(c_0 \le \chi^2_{n-k} \le c_1) = 1 - \alpha$.

A conventional choice is to choose c_0 so that

$$P(\chi^2_{n-k} \le c_0) = \frac{\alpha}{2},$$

and choose c_1 so that

$$P(\chi^2_{n-k} > c_1) = \frac{\alpha}{2}.$$

This can be done easily using tables for the χ^2 distribution.

Alternatively, we could choose c_0, c_1 to minize the length of the confidence interval for σ^2. This will give different values, because the χ^2 distribution is not symmetric.

Hypothesis Tests

Suppose we want to test $H_0 : \sigma^2 = \sigma_0^2$, against either a one-sided or two-sided alternative.

A natural test statistic is:

$$T = \frac{(n-k)s^2}{\sigma_0^2}.$$

Under H_0, this has a χ^2_{n-k} distribution. (The distribution does not depend on β, which is very handy.)

---

[1] We sometimes use the word "inference" to capture testing and confidence intervals.

So, depending on the alternative, we can construct the appropriate critical value. For example, if $H_1 : \sigma^2 > \sigma_0^2$, then we would consider large values of T as evidence against the null hypothesis.

Inference for β

Inference for a single component

Recall that under our normality assumption for y given X:

$$\hat{\beta} \mid X \sim N(\beta, \sigma^2 (X'X)^{-1}).$$

First, suppose that σ^2 is known. Let us consider inference for a particular element of β, say β_j.

Let

$$\sigma_{jj}^2 = \sigma^2 (X'X)_{jj}^{-1},$$

where $(X'X)_{jj}^{-1}$ is the (j,j)th element of $(X'X)^{-1}$, so this is the marginal variance of $\hat{\beta}_j$.

Then

$$\hat{\beta}_j \mid X \sim N(\beta_j, \sigma_{jj}^2).$$

Normalizing, we can write

$$\frac{\hat{\beta}_j - \beta_j}{\sigma_{jj}} \mid X \sim N(0,1),$$

and since the distribution doesn't depend on X, this is also the marginal distribution.

Exercise: Using this result, write down appropriate formulas for confidence intervals, and tests of hypotheses about β_j.

Now, if σ^2 is not known, then we cannot use this approach, because it requires plugging in the true value of σ^2 to get σ_{jj}.

Consider replacing σ^2 with the estimator s^2:

$$T_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{s^2 (X'X)_{jj}^{-1}}}.$$

We can write this as:

$$\begin{aligned} T_j &= \frac{\hat{\beta}_j - \beta_j}{\sqrt{(X'X)_{jj}^{-1}}} \cdot \frac{1}{\sqrt{s^2}} \cdot \frac{\sqrt{\sigma^2}}{\sqrt{\sigma^2}} \cdot \frac{\sqrt{n-k}}{\sqrt{n-k}} \\ &= \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 (X'X)_{jj}^{-1}}} \cdot \frac{1}{\sqrt{(n-k)s^2/\sigma^2}} \cdot \sqrt{n-k}. \end{aligned}$$

The first factor is distributed as $N(0,1)$, and the second factor is independent of the first and distributed χ^2_{n-k}. Recall that if $z \sim N(0,1)$ and $v \sim \chi^2_m$, then

$$\frac{z}{\sqrt{v/m}} \sim t_m,$$

where t_m denotes the t distribution with m degrees of freedom.

Therefore,

$$T_j \sim t_{n-k}.$$

We can then use the tables of the t distribution to construct confidence intervals and hypothesis tests for β_j. Note that if $(n-k)$ is large, the t_{n-k} distribution becomes very close to the standard normal distribution.

The statistic T_j is commonly reported by standard statistical packages, along with the standard error

$$SE_j = \sqrt{s^2 (X'X)^{-1}_{jj}}.$$

We can interpret the standard error as an approximate standard deviation of $\hat{\beta}_j$. When reporting OLS regression results, it is a good idea to report the standard error next to (or below) $\hat{\beta}_j$. A reader can then easily construct relevant confidence intervals and tests using these two numbers. For example, if $n-k$ is large, and we want to test the null hypothesis that $\beta_j = 0$ against a two-sided alternative, the appropriate test statistic would be $|\hat{\beta}_j|/SE_j$, and at the 5% level, we would reject if this quantity was greater than 1.96. Similarly, a 95% confidence interval would be $\hat{\beta}_j \pm (1.96 \cdot SE_j)$ for large $n-k$.

Joint Inference for β

We next consider inference for the entire $k \times 1$ vector β. As before, we start with the case where σ^2 is known.

Recall that

$$\hat{\beta} - \beta \mid X \sim N\left(0, \sigma^2 (X'X)^{-1}\right).$$

by symmetry,

$$\beta - \hat{\beta} \mid X \sim N\left(0, \sigma^2 (X'X)^{-1}\right).$$

We use the following result (Lemma 10.2 in Ruud): suppose that $w \sim N(0, V)$, where V is $m \times m$ and nonsingular. Then

$$w' V^{-1} w \sim \chi^2_m.$$

Using this result, we see that

$$\frac{(\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta})}{\sigma^2} \mid X \sim \chi^2_k.$$

Again, we have that this holds unconditional on X as well.

To get a $(1-\alpha)$ joint confidence region for β, we can choose a value c such that

$$P(\chi^2_k \leq c) = 1 - \alpha.$$

Then our confidence region will be:

$$\left\{\beta \in \mathbb{R}^k \mid \frac{(\beta-\hat{\beta})'(X'X)(\beta-\hat{\beta})}{\sigma^2} \leq c\right\}.$$

This defines an ellipse in $\mathbb{R}^k$.

If we want to test

$$H_0: \quad \beta = \beta_0,$$

vs.

$$H_1: \quad \beta \neq \beta_0,$$

where β_0 now refers to a specific $k \times 1$ constant vector, then we can calculate

$$\frac{(\beta_0-\hat{\beta})'(X'X)(\beta_0-\hat{\beta})}{\sigma^2}$$

and reject the null hypothesis if this value is very large.

If σ^2 is not known, then we need to modify the analysis slightly. Replace σ^2 by s^2, and divide by k to get:

$$\frac{(\beta-\hat{\beta})'(X'X)(\beta-\hat{\beta})/k}{s^2}.$$

We can write this as:

$$\frac{(\beta-\hat{\beta})'(X'X)(\beta-\hat{\beta})/k}{\sigma^2} \cdot \frac{\sigma^2}{s^2(n-k)} \cdot \frac{1/k}{1/(n-k)} \sim \frac{\chi^2_k/k}{\chi^2_{n-k}/(n-k)}.$$

A ratio of two chi-square random variables, each divided by their degrees of freedom, has an F distribution:

$$\frac{\chi^2_k/k}{\chi^2_{n-k}/(n-k)} \sim F_{k,n-k}.$$

So we can use the tables of the F distribution in place of the chi-square distribution to find appropriate joint confidence regions and critical values for tests. For example, if c is such that

$$P(F_{k,n-k} \leq c) = 1-\alpha,$$

then

$$P\left(\frac{(\beta-\hat{\beta})'(X'X)(\beta-\hat{\beta})/k}{s^2} \leq c\right) = 1-\alpha,$$

and a $(1-\alpha)$ confidence region can be formed as:

$$\left\{\beta \in \mathbb{R}^k \mid \frac{(\beta-\hat{\beta})'(X'X)(\beta-\hat{\beta})/k}{s^2} \leq c\right\}.$$

Inference for a linear function of β

Suppose we are interested in inference for a linear combination of the element of β,

$$R\beta = \begin{bmatrix} r_{1,1} & \cdots & r_{1,k} \\ \vdots & & \vdots \\ r_{k-m,1} & \cdots & r_{k-m,k} \end{bmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}.$$

So R is $(k-m) \times k$, and we assume it has full row rank.

Example 1

$$R = (1, 0, 0, \ldots, 0).$$

Then $R\beta = \beta_1$, so this isolates the first element of β.

Example 2

$$R = (0, 1, 1, 1).$$

Then $R\beta = \beta_2 + \beta_3 + \beta_4$.

By the linearity property of the multivariate normal distribution,

$$R\hat{\beta} \mid X \sim N\left(R\beta, R\left[\sigma^2 (X'X)^{-1}\right] R'\right).$$

Therefore,

$$\frac{(R\beta - R\hat{\beta})' \left[R(X'X)^{-1}R'\right]^{-1} (R\beta - R\hat{\beta})}{\sigma^2} \sim \chi^2_{k-m}.$$

We can then use the same reasoning as before to construct confidence regions and tests.

In example 1 above, $R\beta = \beta_1$. So we have an alternative way of doing inference for individual elements of β. It turns out that the two ways are numerically equivalent.