Economics 522A, Spring 2007

Lecture Note 13: Large Sample Inference with OLS

1 Inference for a Single Component of β

In the previous note, we showed that

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\text{d}} N(0, \sigma^2 E[x_i x_i']^{-1}).$$

So, if we are interested in a single component, say β_j,

$$\sqrt{n}(\hat{\beta}_j - \beta_j) \xrightarrow{\text{d}} N(0, \sigma^2 E[x_i x_i']^{-1}_{jj}),$$

where $E[x_i x_i']^{-1}_{jj}$ is the j, j element of the $k \times k$ matrix $E[x_i x_i']^{-1}$. However, we cannot use this result directly for constructing tests and confidence intervals, because the approximate distribution of $\hat{\beta}_j$ depends on σ and $E[x_i x_i']$, which are not known.

Recall that under our assumptions, $s^2 \xrightarrow{\text{p}} \sigma^2$, and

$$\frac{1}{n}\sum_i x_i x_i' \xrightarrow{\text{p}} E[x_i x_i'].$$

A natural estimator of the (normalized) asymptotic variance of $\hat{\beta}_j$ is

$$\hat{V}_j = s^2 \left(\frac{1}{n}\sum_i x_i x_i'\right)^{-1}_{jj}$$

which is consistent:

$$\hat{V}_j \xrightarrow{\text{p}} \sigma^2 E[x_i x_i']^{-1}_{jj}.$$

So consider normalizing the variance of $\hat{\beta}_j$:

$$\sqrt{n}(\hat{\beta}_j - \beta_j)\hat{V}_j^{-1/2} \xrightarrow{\text{d}} N(0,1).$$

Put slightly differently, the quantity

$$\frac{\hat{\beta}_j - \beta_j}{(\hat{V}_j/n)^{1/2}} \overset{\text{a}}{\sim} N(0,1).$$

But

$$\begin{aligned}
\hat{V}_j/n &= s^2 \left(\frac{1}{n}\sum_i x_i x_i'\right)^{-1}_{jj} \frac{1}{n} \\
&= s^2 \left(\sum_{i=1}^n x_i x_i'\right)^{-1}_{jj} \\
&= s^2 (X'X)^{-1}_{jj} \\
&= (SE_j)^2.
\end{aligned}$$

So the quantity

$$\frac{\hat{\beta}_j - \beta_j}{(\hat{V}_j/n)^{1/2}} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{s^2(X'X)^{-1}_{jj}}} = \frac{\hat{\beta}_j - \beta_j}{SE_j},$$

which is the same quantity we used in the normal model, for constructing the t-test and its associated confidence intervals. If we want to test the hypothesis that $\beta_j = c$, for some constant c, we can form a t-statistic:

$$t = \frac{\hat{\beta}_j - c}{SE_j}$$

and compare this to critical values for the standard normal distribution. For example, if the alternative is two-sided, and we want to test the hypothesis at the 5% level, then the asymptotic results imply that

$$Pr(|t| > 1.96) \longrightarrow 0.05$$

under the null hypothesis.

If we want to form a confidence interval, say at the 95% level, our asymptotic distribution result implies that

$$P\left(-1.96 \leq \frac{\hat{\beta}_j - \beta_j}{SE_j} \leq 1.96\right) \longrightarrow 0.95.$$

Equivalently:

$$P\left(\hat{\beta}_j - 1.96SE_j \leq \beta_j \leq \hat{\beta}_j + 1.96SE_j\right) \longrightarrow 0.95.$$

So the confidence interval with endpoints $\hat{\beta}_j \pm 1.96SE_j$ will contain the true β_j approximately 95% percent of the time.

It may be useful to compare this approach to the exact inference for β_j we carried out in the normal model. In both cases, we had a distribution for $\hat{\beta}_j$ that depended on other model parameters. We normalized $\hat{\beta}_j$ by an estimate of its variance.

But in the normal case, using an estimate of the variance changed the distribution of the t-statistic. Fortunately, the t-distribution is relatively easy to work with. In our large-sample argument, using the estimated variance does not alter the asymptotic distribution of the t-statistic — it is still asymptotically normal, because the variance estimator is consistent.

2 Joint Inference for Linear Functions of β

Rather than focus on a single element of β at a time, we can use a similar approach to carry out inference for the entire vector β, and various linear combinations of elements of β.

Let R be a $M \times k$ matrix of constants, with rank M. If $R = I_k$, then $R\beta = \beta$, but other choices for R could be used to isolate subvectors of β or construct other linear combinations of elements of β.

By the delta method,

$$\sqrt{n}(R\hat{\beta} - R\beta) \xrightarrow{\text{d}} N(0, \sigma^2 RE[x_i x_i']^{-1} R').$$

If we want to do inference on $R\beta$, then in practice we will want to replace the variance matrix $\sigma^2 RE[x_i x_i']^{-1} R'$ with an estimated version. Using similar arguments to the previous section,

$$\left[s^2 R(X'X)^{-1} R'\right]^{-1/2} R(\hat{\beta} - \beta) \xrightarrow{\text{d}} N(0, I_M).$$

(The first factor is a matrix, raised to the $-1/2$ power. Recall that matrix powers can be defined using eigenvalue decompositions.)

For carrying out inference, it is also useful to extend this result using the continuous mapping theorem, to get:

$$(\hat{\beta} - \beta)' R' \left[s^2 R(X'X)^{-1} R'\right]^{-1} R(\hat{\beta} - \beta) \xrightarrow{\text{d}} \chi^2_M.$$

So this squared form has a limiting chi-square distribution with M degrees of freedom.

To see how to use this result, consider testing a linear restriction on β:

$$H_0 : \quad R\beta = r,$$

vs.

$$H_1 : \quad R\beta \neq r.$$

Under the null hypothesis $R(\hat{\beta} - \beta) = R\hat{\beta} - r$, so

$$\left(R\hat{\beta} - r\right)' \left[s^2 R(X'X)^{-1} R'\right]^{-1} \left(R\hat{\beta} - r\right) \xrightarrow{\text{d}} \chi^2_M.$$

This has the same form as the approach we used in the normal model. It can also be viewes as a vector version of the simple Wald test we discussed in LN3.

3 Nonlinear Functions of β

Suppose we are interested in some general function $h(\beta)$, where $h : \mathbb{R}^k \mapsto \mathbb{R}^M$.

Assume that h is continuously differentiable in a neighborhood of β, with partial derivative matrix H at β. Then, by the delta method,

$$\sqrt{n}(h(\hat{\beta}) - h(\beta)) \xrightarrow{\text{d}} N(0, \sigma^2 HE[x_i x_i']^{-1} H'),$$

We can then form a Wald-type test statistic in a similar way, this time using the difference between $h(\hat{\beta})$ and $h(\beta)$:

$$(h(\hat{\beta}) - h(\beta))' \left[s^2 H(X'X)^{-1} H'\right]^{-1} (h(\hat{\beta}) - h(\beta)) \xrightarrow{\text{d}} \chi^2_M.$$