Lecture 13

Testing Hypotheses About a Single Linear Combination of the Parameters (See Section 4.4 in Wooldridge.)

Consider the following model:

$$\log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 \, exper + u.$$

Suppose we want to test the null hypothesis that one year at a junior college is worth one year at a university in terms of increasing the salary of a worker. Then

$$H_0 : \beta_1 = \beta_2 .$$

Direct Computation Method

Construct the t-statistic

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2 - 0}{se(\hat{\beta}_1 - \hat{\beta}_2)}$$

Now

$$se(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{\hat{Var}(\hat{\beta}_1) + Var(\hat{\beta}_2) - 2\,\hat{Cov}(\hat{\beta}_1, \hat{\beta}_2)}$$

(we have used $Var(ax + by) = a^2 Var(X) + b^2 Var(y) + 2\,Cov(X, y)$.) Here the ^ (carat) over the Var and Cov represent the sample estimates of the variances of $\hat{\beta}_1$ and $\hat{\beta}_2$ and the $cov(\hat{\beta}_1, \hat{\beta}_2)$.

These numbers, that is $\hat{Var}(\hat{\beta}_1)$, $\hat{Var}(\hat{\beta}_2)$, and $\hat{cov}(\hat{\beta}_1, \hat{\beta}_2)$ are readily available when using the cov option in the PROC REG Model statement, e.g.

```
Proc Reg data = twoyear;
   model lwage = jc univ exper / cov;
```

and the cov option will produce a 4 X 4 variance-covariance matrix like

$$\begin{bmatrix} \hat{Var}(\hat{\beta}_0) & & & \\ \hat{cov}(\hat{\beta}_0, \hat{\beta}_1) & \hat{Var}(\hat{\beta}_1) & & \text{symmetric} \\ \hat{cov}(\hat{\beta}_0, \hat{\beta}_2) & \hat{cov}(\hat{\beta}_1, \hat{\beta}_2) & \hat{Var}(\hat{\beta}_2) & \\ \hat{cov}(\hat{\beta}_0, \hat{\beta}_3) & \hat{cov}(\hat{\beta}_1, \hat{\beta}_3) & \hat{cov}(\hat{\beta}_2, \hat{\beta}_3) & \hat{Var}(\hat{\beta}_3) \end{bmatrix}_{4 \times 4}$$

Then using the values in this matrix you can construct the t-statistic and conduct your test whether it be two tailed or one-tailed. Actually SAS will do this for you if you use a "test statement" as in

```
Proc reg data = twoyear;
  model lwage = jc univ exper
    test jc-univ=0;
```

The resulting output will ~~pr~~ produce an F statistic which can be converted into the t-statistic by the calculation $t = \pm\sqrt{F_{1,N-K}}$ and choosing the right sign of the t by choosing the sign of the t to match the sign of the numerator of the t. Namely choose the negative root if $\hat{\beta}_1 - \hat{\beta}_2 < 0$ and the positive root if $\hat{\beta}_1 - \hat{\beta}_2 > 0$.

Calculating the t-statistic by using a re-parametrized version of the original model

Using the previous example, let $\theta = \beta_1 - \beta_2$. Now $\beta_1 = \theta + \beta_2$. Substituting this into the original model we get the following "reparametrized" model

$$\log(wage) = \beta_0 + (\theta + \beta_2) jc + \beta_2 univ + \beta_3 exper + u$$

$$= \beta_0 + \theta jc + \beta_2 (jc + univ) + \beta_3 exper + u$$

$$= \beta_0 + \theta jc + \beta_2 z + \beta_3 exper + u.$$

Then testing $H_0: \theta = 0$ in this reparametrized equation is equivalent to testing $H_0: \beta_1 - \beta_2 = 0$ in the original equation. Of course in the

reparametrized equation the test of $H_0: \theta = 0$
is automatically calculated by the computer as
$t = \dfrac{\hat{\theta}}{se(\hat{\theta})}$. You don't have to do the tedious
work that is required in the first method.
(Of course, you have to create another data step
to produce the z variable, but that is much
easier to do than getting the variance-covariance
matrix for the OLS estimators and doing the
tedious calculations.)