Econometrics 1
Lecture 4: Finite sample statistical inference
ENSAE 2014/2015

Michael Visser (CREST-ENSAE)

Sampling distributions OLS estimators

Consider again the MLR model

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u \tag{1}$$

Recall from last week's lecture that the OLS estimators are unbiased under assumptions MLR.1-4:

$$E[\hat{\beta}_j] = \beta_j \, \forall j$$

and under MLR.1-5 we have determined the variance of the estimators:

$$Var(\hat{\beta}_j | (x_{1i}, \ldots, x_{ki}) \forall i) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

Recall also that unbiasedness does not mean that the OLS estimate $\hat{\beta}_j$ obtained using a particular sample necessarily coincide with the true value of the parameter, β_j.

Sampling distributions OLS estimators (cntd)

For every sample we will in principle obtain a different estimate of β_j, and only the average estimate (over all samples) will be close to the true value (hypothetical thought experiment since in practice we only have one sample).

In this lecture we wish therefore to answer questions of the following type:

- In what range can we expect β_j to be (with a given probability)?
- How likely is it that β_j equals 0, or $\beta_j = \beta_l$? etc ...

We will need to know the full sampling distribution of $\hat{\beta}_j$ to answer these questions. Knowing the first two moments of $\hat{\beta}_j$ (expectation and variance) is not enough to perform statistical inference.

Sampling distributions OLS estimators (cntd)

Since (see slides of last week)

$$\hat{\beta}_j = \beta_j + \frac{\sum_{i=1}^{n} \hat{r}_{ij} u_i}{\sum_{i=1}^{n} \hat{r}_{ij}^2}$$

the *sampling* distribution of $\hat{\beta}_j$ depends on the *population* distribution of the errors once we condition on $x_{ij}\ \forall i, j$.

To make the sampling distribution tractable we assume that the error term follows a normal distribution in the population.

MLR.6 - Normality of error term

The population error u is independent of $x_1, ..., x_k$, and normally distributed with zero mean and variance σ^2:

$$u | x_1, ..., x_k \sim \mathcal{N}(0, \sigma^2)$$

Sampling distributions OLS estimators (cntd)

Remarks:

- MLR.6 is a strong assumption. Since u is independent of $x_1, ..., x_k$ we have that $E[u|x_1, ..., x_k] = E[u] = 0$ and $Var[u|x_1, ..., x_k] = Var[u] = \sigma^2$. MLR.6 therefore implies MLR.3 (zero conditional mean) and MLR.5 (homoscedasticity).
- MLR.1-6 imply that conditionally on the regressors the dependent variable has a normal distribution:

$$y|x_1, ..., x_k \sim \mathcal{N}(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k, \sigma^2)$$

Sampling distributions OLS estimators (cntd)

- Whether conditional normality of the dependent variable can be assumed is essentially an empirical matter. When the empirical distribution of $y|x$ is very different from the normal, a transformation of the model may sometimes lead to a better approximation. In some contexts the conditional normality assumption is inappropriate (e.g., discrete dependent variable, dependent variable with much mass at 0, etc.).
- Under MLR.1-6 the OLS estimators are the *minimum variance unbiased estimators*: they have the smallest variance among *all* unbiased estimators (i.e. not only those linear in y).

Sampling distributions OLS estimators (cntd)

The next theorem establishes that the OLS estimator follows a normal distribution.

Normal sampling distribution of OLS estimator

Under MLR.1-6

$$\hat{\beta}_j | X \sim \mathcal{N}(\beta_j, var(\hat{\beta}_j | X))$$

where

$$Var(\hat{\beta}_j | X) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

$X = ((x_{i1}, ..., x_{ik}), \forall i = 1, ..., n)$, $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$ and R_j^2 is the R-squared from a regression of x_j on the other regressors (and an intercept).

Sampling distributions OLS estimators (cntd)

Remarks:

- $\hat{\beta}_j|X$ is normally distributed because $\hat{\beta}_j$ is a weighted sum of i.i.d. normal random variables (the weights $w_{ij} = \hat{r}_{ij} / \sum_{i=1}^{n} \hat{r}_{ij}^2$ are non-random given X and the u_i are i.i.d. $\mathcal{N}(0, \sigma^2)$):

$$\hat{\beta}_j = \beta_j + \sum_{i=1}^{n} w_{ij} u_i.$$

- The normality of the estimators has a number of implications:
  - Any linear combination of the $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ is also normally distributed
  - Any subset of the $\hat{\beta}_j$ has a *joint* normal distribution
  - Letting $sd(\hat{\beta}_j|X) = \sqrt{Var(\hat{\beta}_j|X)}$ be the conditional standard deviation of $\hat{\beta}_j$, we have conditionally on X:

$$(\hat{\beta}_j - \beta_j)/sd(\hat{\beta}_j|X) \sim \mathcal{N}(0, 1) \tag{2}$$

t Distribution

The following result is crucial in constructing hypotheses tests and confidence intervals.

t Distribution for standardized OLS estimators

$$(\hat{\beta}_j - \beta_j)/se(\hat{\beta}_j|X) \sim t_{n-k-1} \tag{3}$$

where $k+1$ is the number of unknown parameters in the population model (1), n the sample size, the standard error $se(\hat{\beta}_j|X)$ is the estimate of the standard deviation $sd(\hat{\beta}_j|X)$:

$$se(\hat{\beta}_j|X) = \sqrt{\widehat{Var}(\hat{\beta}_j|X)} = \sqrt{\frac{\hat{\sigma}^2}{SST_j(1-R_j^2)}}$$

and $\hat{\sigma} = (n-k-1)^{-1/2}\sqrt{\sum_{i=1}^{n} \hat{u}_i^2}$

t Distribution (cntd)

Remark:
The standardized estimator (3) follows a t-distribution with $n-k-1$ degrees of freedom (and not a $N(0,1)$ as in (2)) because the unknown standard deviation $sd(\hat{\beta}_j|X)$ has been replaced by the standard error $se(\hat{\beta}_j|X)$, i.e., the error variance σ^2 has been replaced by its unbiased estimate $\hat{\sigma}^2 = (n-k-1)^{-1}\sum_{i=1}^{n}\hat{u}_i^2$.

Hypothesis on a single parameter

We now turn to the issue of hypothesis testing in the regression model (1) which satisfies MLR.1-6.

The β_j's are unknown features of the population; we know nothing about them with certainty but:

- We can formulate a hypothesis about their value
- Construct a test statistic with a known distribution under the maintained hypothesis
- Reject our hypothesis if the value of the test statistic is "too unlikely to happen under the maintained hypothesis"

We start discussing tests that involve a single parameter.

Null and alternative hypothesis

We first need to formulate a *null hypothesis*, denoted $\mathcal{H}_0$. For example

"x_j has no effect on y once the other x's are controlled for".

This is an important type of null hypothesis in practice. It can be written formally as

$$\mathcal{H}_0 : \beta_j = 0 \tag{4}$$

Next we need to confront $\mathcal{H}_0$ with an *alternative hypothesis*, denoted $\mathcal{H}_1$. For example

$$\mathcal{H}_1 : \beta_j > 0 \tag{5}$$

or

$$\mathcal{H}_1 : \beta_j < 0 \tag{6}$$

These are examples of so-called one-sided alternatives. An example of a two-sided alternative would be

$$\mathcal{H}_1 : \beta_j \neq 0 \tag{7}$$

One-sided alternatives

To test the null hypothesis $\mathcal{H}_0 : \beta_j = 0$ against the one-sided alternative $\mathcal{H}_1 : \beta_j > 0$ we use the so-called t statistic:

$$t_{\hat{\beta}_j} \equiv \hat{\beta}_j / se(\hat{\beta}_j) \tag{8}$$

- under $\mathcal{H}_0$, $t_{\hat{\beta}_j}$ follows a t distribution and has zero mean
- under $\mathcal{H}_1$, $t_{\hat{\beta}_j}$ has a positive mean

Suppose for example that $\beta_j = 10$. It follows from (3) that $(\hat{\beta}_j - 10)/se(\hat{\beta}_j) \sim t_{n-k-1}$. The statistic (8) can then be written as

$$t_{\hat{\beta}_j} = \frac{(\hat{\beta}_j - 10)}{se(\hat{\beta}_j)} + \frac{10}{se(\hat{\beta}_j)}$$

The conditional expectation of this statistic equals

$$0 + 10/se(\hat{\beta}_j) = 10/se(\hat{\beta}_j)$$

which is positive.

One-sided alternatives (cntd)

So it makes sense to reject $\mathcal{H}_0$ if $t_{\hat{\beta}_j}$ is large, and accept the null otherwise.

To define more precisely the rejection rule, it is required to fix the probability of a Type I error, or significance level. This is the probability of rejecting the null when it is actually true.

Let α denote the significance level, and suppose we choose $\alpha = 5\%$ (the most common value for α; sometimes $\alpha = 1\%$ or $\alpha = 10\%$ is chosen). The rejection rule is that $\mathcal{H}_0$ is rejected in favor of $\mathcal{H}_1$ at the 5% significance level if

$$t_{\hat{\beta}_j} > c$$

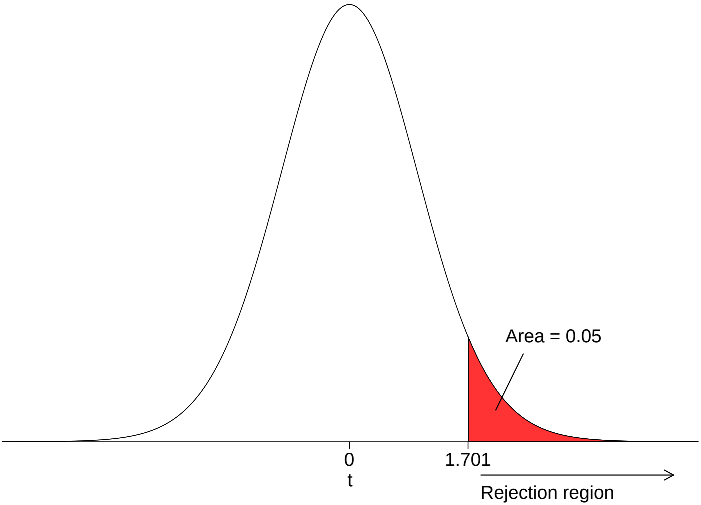
where c is the 5% critical value for a one-sided test, i.e., it is the solution of

$$5\% = Pr(\text{reject } H_0 | H_0 \text{ is true}) = Pr(t_{n-k-1} > c)$$

One-sided alternatives (cntd)

Thus c is chosen such that the area in the right tail of the t distribution equals 5%. In other words, c is the 95-th percentile in the t distribution with $n - k - 1$ degrees of freedom. The next figure illustrates this for the case $n - k - 1 = 28$ (find a table with critical values of a t distribution and check that $c = 1.701$. Check also that when $\alpha = 1\%$, $c = 2.467$, and when $\alpha = 10\%$, $c = 1.313$.)

5% rejection rule for $\mathcal{H}_1 : \beta_j > 0$ (df=28)



One-sided alternatives (cntd)

The procedure is similar when we wish to test the null hypothesis $\mathcal{H}_0 : \beta_j = 0$ against the one-sided alternative $\mathcal{H}_1 : \beta_j < 0$.

The rejection rule is that $\mathcal{H}_0$ is rejected in favor of $\mathcal{H}_1$ at the 5% significance level if

$$t_{\hat{\beta}_j} < c$$

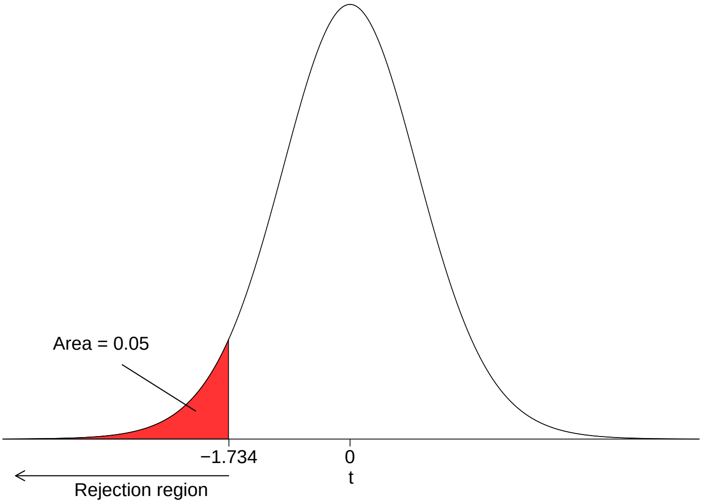
where c is the 5% critical value for a one-sided test, i.e., it is the solution of

$$5\% = Pr(\text{reject } H_0 | H_0 \text{ is true}) = Pr(t_{n-k-1} < c)$$

Thus c is chosen such that the area in the left tail of the t distribution equals 5%: c is the 5-th percentile in the t distribution with $n - k - 1$ degrees of freedom. The next figure illustrates this

for the case $n - k - 1 = 18$ (check that when $\alpha = 5\%$, $c = -1.734$).

5% rejection rule for $\mathcal{H}_1 : \beta_j < 0$ (df=18)



Example: Exam performance

```
. reg exam presence moy

      Source |       SS       df       MS              Number of obs =      85
-------------+------------------------------           F(  2,    82) =   13.47
       Model |  253.446064     2  126.723032           Prob > F      =  0.0000
    Residual |  771.542171    82  9.40905087           R-squared     =  0.2473
-------------+------------------------------           Adj R-squared =  0.2289
       Total |  1024.98824    84  12.2022409           Root MSE      =  3.0674

                                      +-------+
--------------------------------------|-------|--------------------------------
        exam |      Coef.   Std. Err. |     t | P>|t|     [95% Conf. Interval]
-------------+------------------------|-------|--------------------------------
    presence |   .0388285   .1159276  |  0.33 | 0.739    -.1917883    .2694454
       moy1a |   1.470245   .3067414  |  4.79 | 0.000     .8600386    2.080451
       _cons |  -7.096716    3.92975  | -1.81 | 0.075    -14.91424    .7208074
--------------------------------------|-------|--------------------------------
                                      +-------+

. di invttail(82,0.05)
1.6636492
```

The last Stata command gives the 95-th percentile of a t-distribution with 82 degrees of freedom. So the 5% critical value for the one-sided test $\mathcal{H}_0 : \beta_j = 0$ vs $\mathcal{H}_1 : \beta_j < 0$ is -1.66.

Two-sided alternatives

It is common to test $\mathcal{H}_0$ against the *two-sided* alternative

$$\mathcal{H}_1 : \beta_j \neq 0$$

- this is a natural alternative hypothesis when there is no prior knowledge about the sign
- it also keeps us honest: not changing the alternative after seeing the sign of $\hat{\beta}_j$ (you need to state the alternative before looking at the the data)!

Now the rejection rule is that $\mathcal{H}_0$ is rejected in favor of $\mathcal{H}_1$ if

$$|t_{\hat{\beta}_j}| > c$$

At significance level α we now choose c such that the probability mass in each tail of the distribution of $t_{\hat{\beta}_j}$ equals $\alpha/2$.

Two-sided alternatives (cntd)

For $\alpha = 5\%$ we get

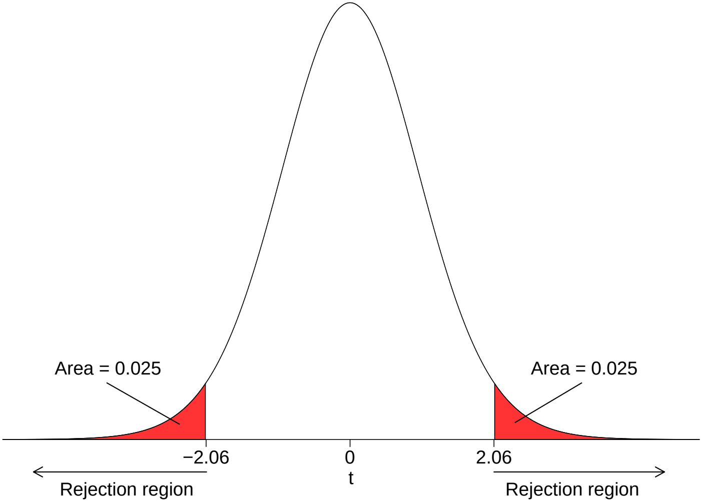
$$5\% = Pr(\text{reject } H_0 | H_0 \text{ is true}) = Pr(|t_{n-k-1}| > c)$$

Thus c is chosen such that the area in each tail of the t distribution equals 2.5%: c is the 97.5-th percentile in the t distribution with $n-k-1$ degrees of freedom. Terminology :

- if H_0 is rejected we say "x_j is statistically significant/statistically different from zero, at the 5% percent level"
- if H_0 is not rejected we say "x_j is statistically insignificant at the 5% level"

The next figure illustrates the rejection rule for the case $n-k-1 = 25$ (check that when $\alpha = 5\%$, $c = 2.06$).

5% rejection rule for $\mathcal{H}_1 : \beta_j \neq 0$ (df=25)



Other hypotheses about β_j

We may want to test other null hypotheses such as

$$\mathcal{H}_0 : \beta_j = a_j \tag{9}$$

where a_j is the hypothesized value. The above testing theory can easily be adapted to this case. Consider the general t statistic

$$t = \frac{(\hat{\beta}_j - a_j)}{se(\hat{\beta}_j)} \tag{10}$$

The statistic (10) can be used to test (9) against one-sided or two-sided alternatives. The calculation of critical values proceeds in *exactly* the same way as above. The only thing that changes is how we calculate the test statistic.

Other hypotheses about β_j (cntd)

Consider a simple linear regression model relating *crime* (annual number of crimes on a college campus) to *enroll* (student enrollment at the college)

$$log(crime) = \beta_0 + \beta_1 log(enroll) + u$$

The OLS estimates (standard errors) obtained using a random sample of 97 colleges and universities in the U.S. in 1992 (source: FBI's *Uniform Crime Reports*) are: $\hat{\beta}_0 = -6.63$ (1.03), and $\hat{\beta}_1 = 1.27$ (0.11) (the estimate of the slope suggests that increasing enrollment by 1% augments the number of crimes by 1.27%). Let

us test $H_0 : \beta_1 = 1$ against $H_1 : \beta_1 > 1$. The statistic (10) equals (1.27-1)/0.11=2.45. The one-sided 5% critical value for a t

distribution with 97-1-1=95 degrees of freedom is approximately 1.66. The outcome of the statistic being larger than the critical value, we reject H_0 in favor of H_1. The 1% critical value roughly equals 2.37, so the null is rejected even at the 1% level.

Computing p-values

In describing empirical results, it is useful to report not only the t statistics but also the so-called p-values of t tests. By definition, the p-value of a t test is the *smallest significance level* at which the null hypothesis would be rejected given the outcome of the t statistic. The p-value summarizes the strength with which the null can be rejected. For example, suppose we desire to test that β_j equals zero against the two-sided alternative that it differs from zero. Suppose there are 40 degrees of freedom, and the outcome of the t statistic is 1.85. Since the two-tailed 5% critical value is 2.02 and the 10% critical value 1.68, we would reject the null at the 10% significance level but not at the 5% level. So necessarily the p-value is between 5 and 10%. Its precise value is obtained by calculating the probability that the absolute value of a student random variable with 40 degrees of freedom exceeds 1.85:

$$p-\text{value} = Pr(|t_{40}| > 1.85) = 2Pr(t_{40} > 1.85) = 2(0.0359) = 0.07.$$

Example: Exam performance

```
. reg exam presence moy

      Source |       SS       df       MS              Number of obs =      85
-------------+------------------------------           F(  2,    82) =   13.47
       Model |  253.446064     2  126.723032           Prob > F      =  0.0000
    Residual |  771.542171    82  9.40905087           R-squared     =  0.2473
-------------+------------------------------           Adj R-squared =  0.2289
       Total |  1024.98824    84  12.2022409           Root MSE      =  3.0674

                                                +-------+
------------------------------------------------|-------|------------------------
        exam |      Coef.   Std. Err.      t    | P>|t| |     [95% Conf. Interval]
-------------+----------------------------------|-------|------------------------
    presence |   .0388285   .1159276     0.33   | 0.739 |   -.1917883    .2694454
       moy1a |   1.470245   .3067414     4.79   | 0.000 |    .8600386    2.080451
       _cons |  -7.096716    3.92975    -1.81   | 0.075 |   -14.91424    .7208074
------------------------------------------------|-------|------------------------
                                                +-------+
```

Confidence intervals

In empirical studies it is also common to report a *confidence interval.* A *confidence interval* = range of likely values for the population parameter. Using the fact that $(\hat{\beta}_j - \beta_j)/se(\hat{\beta}_j) \sim t_{n-k-1}$ and after some simple calculations, we get a $(1-\alpha)$ confidence interval for the population parameter β_j:

$$\left(\hat{\beta}_j - c \cdot se(\hat{\beta}_j),\, \hat{\beta}_j + c \cdot se(\hat{\beta}_j)\right)$$

where c is the $(1-\frac{\alpha}{2})-$th percentile in a t_{n-k-1} distribution. A CI has the following interpretation: suppose many random samples are available, and that a confidence interval is calculated for each sample, then a fraction$(1-\alpha)$ of the intervals would cover the population value β_j.

Confidence intervals (cntd)

Remarks:

- For a 95% confidence interval c would be the 97.5th percentile in a t_{n-k-1} distribution. Rules of thumb for 95% CI's:
  - $c \approx 2$ if $df > 50$
  - $c = 1.96$ if $df > 120$ (the t distribution gets close to a standard normal)
  - Use student table to determine c for small degrees of freedom
- CI is as good as the underlying assumptions: are estimates unbiased? is $se(\hat{\beta}_j)$ correct (heteroscedasticity? normality of error term?)?
- Regression packages typically report CI's
- With a CI it is easy to perform *two-tailed* tests: Reject the null $H_0 : \beta_j = a_j$ in favor of $H_1 : \beta_j \neq a_j$ when a_j is not in the CI

Confidence intervals (cntd)

In the crime/enrollment example, the 95% confidence interval for β_1 is:

$$[1.27 - 1.99 \cdot (0.11); 1.27 + 1.99 \cdot (0.11)] = [1.05; 1.49].$$

where we used the fact that the 97.5-th percentile in a t_{95} distribution approximately equals 1.99. Since the value “1” is not in the interval, we reject H_0 at the 5% level in favor of $H_1 : \beta_1 \neq 1$.

Example: Exam performance

```
. reg exam presence moy

      Source |       SS       df       MS              Number of obs =      85
-------------+------------------------------           F(  2,    82) =   13.47
       Model |  253.446064     2  126.723032           Prob > F      =  0.0000
    Residual |  771.542171    82  9.40905087           R-squared     =  0.2473
-------------+------------------------------           Adj R-squared =  0.2289
       Total |  1024.98824    84  12.2022409           Root MSE      =  3.0674

                                                       +-----------------------+
-------------------------------------------------------|---------------------- |
        exam |      Coef.   Std. Err.      t    P>|t|  |  [95% Conf. Interval] |
-------------+-----------------------------------------|---------------------- |
    presence |   .0388285   .1159276     0.33   0.739  | -.1917883    .2694454 |
       moy1a |   1.470245   .3067414     4.79   0.000  |  .8600386    2.080451 |
       _cons |  -7.096716    3.92975    -1.81   0.075  | -14.91424    .7208074 |
-------------------------------------------------------|---------------------- |
                                                       +-----------------------+

. scalar c = invttail(82,0.025)
. di _b[presence] - c*_se[presence]
-.19178832
. di _b[presence] + c*_se[presence]
.26944539
```

Test of linear combination of parameters

In applications, the hypothesis of interest may involve one or several combinations of the parameters. The null hypothesis can be of the form:

$$\mathcal{H}_0 : \beta_j = \beta_k$$

we can rewrite this as

$$\mathcal{H}_0 : \beta_j - \beta_k = 0$$

and use the approach above using the following t statistic

$$t = \frac{\hat{\beta}_j - \hat{\beta}_k}{se(\hat{\beta}_j - \hat{\beta}_k)} \tag{11}$$

The complication is that we need an estimate of the covariance between $\hat{\beta}_j$ and $\hat{\beta}_k$ since

$$se(\hat{\beta}_j - \hat{\beta}_k) = \sqrt{\widehat{var}(\hat{\beta}_j) + \widehat{var}(\hat{\beta}_k) - 2\widehat{cov}(\hat{\beta}_j, \hat{\beta}_k)}$$

Test of linear combination of parameters (cntd)

We showed last week that

$$\hat{\beta} = (X'X)^{-1}X'y$$

and

$$\widehat{Var}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1} \tag{12}$$

where

$$\hat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i{}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2$$

The standard error $se(\hat{\beta}_j - \hat{\beta}_k)$ can now be calculated using (12): $\widehat{var}(\hat{\beta}_j)$ and $\widehat{var}(\hat{\beta}_k)$ (last week we also gave the explicit scalar expressions of these estimated variances) are the j-th and k-th elements of the diagonal of $\widehat{Var}(\hat{\beta})$, and $\widehat{cov}(\hat{\beta}_j, \hat{\beta}_k)$ is the (j, k)-th element of $\widehat{Var}(\hat{\beta})$.

Test of linear combination of parameters (cntd)

Remarks:

- Most econometric/statistical software packages calculate $se(\hat{\beta}_j - \hat{\beta}_k)$ for you, which in turn allows you to determine the statistic (11).
- The test procedure exploits the fact that (11) follows a t_{n-k-1} distribution under the null (thus with mean zero), but has a positive (resp. negative) mean if $\beta_j > \beta_k$ (resp. $\beta_j < \beta_k$). The testing procedures (rejection rule, calculation of critical value, etc.) are similar to the procedures described above.
- There is, however, a simpler way to perform the test ...

Test of linear combination of parameters (cntd)

Suppose we want to test

$$\mathcal{H}_0 : \beta_1 = \beta_2$$

in the population model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

Denoting $\theta = \beta_1 - \beta_2$ we can rewrite it as follows:

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \\ &= \beta_0 + (\theta + \beta_2) x_1 + \beta_2 x_2 + u \\ &= \beta_0 + \theta x_1 + \beta_2 (x_1 + x_2) + u \end{aligned}$$

and see that

$$\mathcal{H}_0 : \beta_1 = \beta_2 \iff \mathcal{H}_0 : \theta = 0$$

Test of linear combination of parameters (cntd)

Remarks:

- In the testing procedure we still need to be explicit about H1 (one-sided/two-sided)
- This approach works (also with more than 2 variables) as long as we have just one restriction
- Actually software packages such as Stata does this easily for you by typing the command “test x1=x2” (after the regression)

Test of multiple linear restrictions on parameters

Suppose we want to test whether a group of q variables has no effect on the outcome. More precisely, in the model (1) we wish to test the null hypothesis:

$$\mathcal{H}_0 : \beta_{k-q+1} = 0, ..., \beta_k = 0. \tag{13}$$

We consider two-sided alternatives

$$\mathcal{H}_1 : \mathcal{H}_0 \text{ is not true}$$

The null puts q exclusion restrictions on the parameters. Note that it is not restrictive to impose these restrictions on the q *last* parameters since the order in which the k variables appear in model (1) is arbitrary.

Under the null, model (1) becomes

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_{k-q} x_{k-q} + u$$

This version of the model is called the restricted model, and (1) the unrestricted model.

Test of multiple linear restrictions on parameters (cntd)

The statistic used to test (13) is the so-called F statistics:

F test

$$F \equiv \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} \tag{14}$$

where SSR_{ur} is the sum of squared residuals from the unrestricted model, and SSR_r the sum of squared residuals from the restricted model.

Since the OLS method consists in minimizing the sum of squared residuals, the SSR from a given model increases (or remains unchanged) when regressors are dropped from the model: $SSR_r - SSR_{ur} \geq 0$.

The larger the difference between SSR_r and SSR_{ur} the stronger the evidence against H_0. So it seems natural to reject the null hypothesis if the outcome of the statistic (14) is "large".

Test of multiple linear restrictions on parameters (cntd)

In defining the rejection rule we use that under the null the statistic (14) follows a F distribution with $(q, n-k-1)$ degrees of freedom.

Rejection rule: reject H_0 in favor of H_1 at the significance level α if

$$F > c$$

where c is the $(1-100\alpha)$-th percentile in the F distribution with $(q, n-k-1)$ degrees of freedom, i.e., it is the solution of: $\alpha = Pr(F_{q,n-k-1} > c)$.

Remarks:

- Using that $SSR_r = SST(1-R_r^2)$ and $SSR_{ur} = SST(1-R_{ur}^2)$, the statistic (14) can be rewritten as

$$F \equiv \frac{(R_{ur}^2 - R_r^2)/q}{(1-R_{ur}^2)/(n-k-1)}. \tag{15}$$

  This is sometimes called the R-squared form of the F statistic.

Test of multiple linear restrictions on parameters (cntd)

- When the null hypothesis (13) is $H_0 : \beta_1 = 0, ..., \beta_k = 0$ (all slope parameters are zero), the R-squared form of the F statistic is
$$F = \frac{R^2_{ur}/k}{(1 - R^2_{ur})/(n - k - 1)}$$
since $R^2_r = 0$ in a model where y is just regressed on a constant, i.e., without regressors.
- The F test can easily be adapted to the case where the null hypothesis of interest is of the form $H_0 : \beta_{k-q+1} = a_{k-q+1}, ..., \beta_k = a_k$, where $a_{k-q+1}, ..., a_k$ are the hypothesized values. For example consider the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u$, and we want to test $\mathcal{H}_0 : \beta_1 = c$, $\beta_2 = 0$, $\beta_3 = 0$, $\beta_4 = 0$. We therefore need estimate the restricted model $y = \beta_0 + c x_1 + u$. To do so we transform the dependent variable: $y - c x_1 = \beta_0 + u$. Using the SSR_r from this regression we can calculate the F statistic.

Example: Birth weight

We we are interested in the relationship between the infant birth weight (a measure of infant health) in ounces (1 ounce=28.35 grams), denoted *bwght*, and a number of regressors: *cigs* (average number of cigarettes smoked by mother per day during pregnancy), *faminc* (annual family income in dollars), *parity* (birth order of the child), *motheduc* (years of schooling of the mother), and *fatheduc* (years of schooling of father).

The estimated birth weight equation is

$$
\begin{aligned}
(\widehat{bwght}) = \underset{(3.73)}{114.52} \underset{(0.11)}{-0.60}\, cigs + \underset{(0.66)}{1.79}\, parity + \underset{(0.04)}{0.06}\, faminc \\
\underset{(0.32)}{-0.37}\, motheduc + \underset{(0.28)}{0.47}\, fatheduc. \\
n = 1191, R^2 = 0.039, SSR = 464041.13.
\end{aligned}
$$

Example: Birth weight (cntd)

Let us test the null that parents' education does not have an effect on the weight of the infant:

$$H_0 : \beta_4 = 0, \beta_5 = 0$$

The restricted birth weight equation is

$$\widehat{(bwght)} = \underset{(1.66)}{115.47} \underset{(0.11)}{-0.60}\, cigs + \underset{(0.66)}{1.83}\, parity + \underset{(0.03)}{0.07}\, faminc.$$
$$n = 1191, R^2 = 0.036, SSR = 465166.79.$$

We have $n - k - 1 = 1191 - 6 = 1185$, and $q = 2$. The 5% critical value for the $F_{2,1185}$ distribution is $c = 3$. The F statistic (15) equals 1.42, so we cannot reject the null: *motheduc* and *fatheduc* are jointly insignificant in the birth weight equation.

I understand/can apply...

- The OLS estimators are normally distributed under assumptions MLR.1-6
- We test hypotheses about <u>unknown</u> population parameters
    - null hypotheses
    - one- or two-sided alternatives
    - p-values, significance levels, critical values
- The t statistic
    - used to test hypotheses about a *single* parameter
    - under MLR.1-6, the t statistics have t distributions under $\mathcal{H}_0$
- Confidence intervals
- The F statistic
    - used to test *multiple* exclusion restrictions,
    - the numerator df is the number of restrictions being tested, while the denominator df is the degrees of freedom in the unrestricted model
    - The alternative for F testing is two-sided