

LECTURE 5

HYPOTHESIS TESTING

Basic concepts

In this lecture, we continue to discuss the normal classical linear regression model defined by Assumptions (A1)–(A5). Let $\theta \in \Theta \subset \mathbb{R}^d$ be a parameter of interest. Some examples of θ include:

- The coefficient of one of the regressors: $\theta = \beta_1$, $d = 1$, $\Theta = \mathbb{R}$.
- A vector of coefficients: $\theta = (\beta_1, \dots, \beta_l)'$, $d = l$, $\Theta = \mathbb{R}^l$.
- The variance of errors: $\theta = \sigma^2$, $d = 1$, $\Theta = \mathbb{R}_{++}$.

A statistical hypothesis is an assertion about θ . Usually, we have two competing hypotheses, and we want to draw a conclusion, based on the data, as to which of the hypotheses is true. Let $\Theta_0 \subset \Theta$ and $\Theta_1 \subset \Theta$ such that $\Theta_0 \cap \Theta_1 = \emptyset$ and $\Theta_0 \cup \Theta_1 = \Theta$. The two competing hypotheses are:

- Null hypothesis $H_0 : \theta \in \Theta_0$. This is a hypothesis that is held as true, unless the data provide *sufficient* evidence against it.
- Alternative hypothesis $H_1 : \theta \in \Theta_1$. This is a hypothesis against which the null is tested. It is held to be true if the null is found false.

The subsets Θ_0 and Θ_1 are chosen by the econometrician, and therefore are *known*. Usually, the econometrician has to carry the burden of proof, so the case that they are interested in is stated as H_1 .

The two hypotheses, H_0 and H_1 , must be *disjoint*. Their union defines the *maintained* hypothesis, that is, the space of values that θ can take. For example, when $\Theta = \mathbb{R}$, one may consider $\Theta_0 = \{0\}$ and $\Theta_1 = \mathbb{R} \setminus \{0\}$. Another example is $\Theta_0 = (-\infty, 0]$ and $\Theta_1 = (0, \infty)$.

When Θ_0 has exactly one element (Θ_0 is a singleton), we say that $H_0 : \theta \in \Theta_0$ is a *simple* hypothesis. Otherwise, we say that H_0 is a *composite* hypothesis. Similarly, $H_1 : \theta \in \Theta_1$ can be simple or composite depending on whether Θ_1 is a singleton or not.

The econometrician has to choose between H_0 and H_1 . The *decision rule* that leads the econometrician to *reject or fail to reject* H_0 is based on a *test statistic*, which is a function of the data (X and Y in the case of a regression model). Let $S \in \mathcal{S}$ denote a test statistic, where \mathcal{S} is the range of its values. A decision rule is defined by a partition of \mathcal{S} into an acceptance region \mathcal{A} and a rejection (critical) region \mathcal{R} . The acceptance and rejection regions must be disjoint ($\mathcal{A} \cap \mathcal{R} = \emptyset$), and their union must be equal to the range of possible values for S ($\mathcal{A} \cup \mathcal{R} = \mathcal{S}$). One rejects H_0 when the test statistic falls into the *rejection region*: $S \in \mathcal{R}$. Thus, tests can be described by their decision rules: Reject H_0 when $S \in \mathcal{R}$.

There are two types of errors that the econometrician can make:

- Type I error is the error of rejecting H_0 when H_0 is true.
- Type II error is the error of failing to reject H_0 when H_1 is true.

The probabilities of Type I and II errors can be described using the so-called *power function*. Consider a test based on S that rejects H_0 when $S \in \mathcal{R}$. The power function of this test is defined as:

$$\pi(\theta) = \Pr(S \in \mathcal{R} \mid \theta),$$

where the conditioning on θ means that the probability is computed given that the true value of the parameter is θ . Thus, the power function of a test gives the probability of rejecting H_0 for every possible value of θ . The largest probability of Type I error (rejecting H_0 when it is true) is

$$\sup_{\theta \in \Theta_0} \pi(\theta) = \sup_{\theta \in \Theta_0} \Pr(S \in \mathcal{R} \mid \theta). \tag{1}$$

The expression above is also called the *size* of a test. When H_0 is simple, that is, $\Theta_0 = \{\theta_0\}$, the size can be computed simply as $\pi(\theta_0) = \Pr(S \in \mathcal{R} \mid \theta_0)$.

The probability of Type II error (failing to reject H_0 when it is false) is:

$$1 - \pi(\theta) = 1 - \Pr(S \in \mathcal{R} \mid \theta) \quad \text{for } \theta \in \Theta_1. \quad (2)$$

Typically, Θ_1 has many elements, and therefore the probability of Type II error depends on the true value θ . One would like the probabilities of both Type I and Type II errors to be as small as possible, but unfortunately, they are inversely related, as is apparent from (1) and (2). To reduce the probability of Type I error (falsely rejecting H_0), one should make \mathcal{R} smaller. This, however, will increase the probability of Type II error.

By convention, a *valid* test must control the size (probability of Type I error). This is consistent with the idea that the econometrician must carry the burden of proof (recall that the econometrician must state their preferred hypothesis as H_1).

Definition. A test with power function $\pi(\theta)$ is said to be a level α test if $\sup_{\theta \in \Theta_0} \pi(\theta) \leq \alpha$. We say it is a size α test if $\sup_{\theta \in \Theta_0} \pi(\theta) = \alpha$.

Every size α test is a level α test. We consider a test to be valid if it is a level α test for some pre-chosen $\alpha \in (0, 1)$, where α is called the *significance level* of a test. Typically, the significance level is chosen to be a small number close to zero: for example, $\alpha = 0.01, 0.05, 0.10$.

The following are the steps of hypothesis testing:

1. Specify H_0 and H_1 .
2. Choose the significance level α .
3. Define a decision rule (a test statistic and a rejection region) so that the resulting test is a level α test.
4. Perform the test.

The decision depends on the significance level: it is easier to reject the null for larger values of α , since they correspond to larger rejection regions. Given data, the smallest significance level at which the null can be rejected is called the *p-value*. Instead of reporting test outcomes (reject or fail to reject) for some specific α , it is also common to report *p-values*:

1. Specify H_0 and H_1 .
2. Define a test.
3. Compute the *p-value*.
4. H_0 is rejected for all values of α that are *greater* than the *p-value*.

The *power of a test* with power function $\pi(\theta)$ is defined as

$$\pi(\theta) \quad \text{for } \theta \in \Theta_1.$$

Given two level α tests, we should prefer a more powerful test. We say that a level α test with power function $\pi_1(\theta)$ is *uniformly* more powerful than a level α test with power function $\pi_2(\theta)$ if $\pi_1(\theta) \geq \pi_2(\theta)$ for all $\theta \in \Theta_1$. As will be apparent from the next section, tests based on estimators with smaller variances typically result in uniformly more powerful tests.

Testing a hypothesis about a single coefficient

Consider the partitioned regression discussed in Lecture 4:

$$Y = \beta_1 X_1 + X_2 \beta_2 + U,$$

where X_1 is the $n \times 1$ vector of the observations of the first regressor. Assume that the disturbance variance σ^2 is known. Let $\hat{\beta}_1$ be the LS estimator of β_1 . Suppose we want to test

$$\begin{aligned} H_0 : \beta_1 &= \beta_{1,0}, \\ H_1 : \beta_1 &\neq \beta_{1,0}. \end{aligned} \tag{3}$$

Confidence intervals and hypothesis testing are closely related. In fact, a decision rule for an α -level test can be based on the $CI_{1-\alpha}$. The $1 - \alpha$ level confidence interval for β_1 is

$$CI_{1-\alpha} = \left[\hat{\beta}_1 - z_{1-\alpha/2} \sqrt{\sigma^2 / (X_1^\top M_2 X_1)}, \hat{\beta}_1 + z_{1-\alpha/2} \sqrt{\sigma^2 / (X_1^\top M_2 X_1)} \right].$$

Consider the following test:

$$\text{Reject } H_0 \text{ if } \beta_{1,0} \notin CI_{1-\alpha}.$$

The critical region in this case is given by the complement of the $CI_{1-\alpha}$. Thus, we reject if

$$\begin{aligned} \beta_{1,0} &< \hat{\beta}_1 - z_{1-\alpha/2} \sqrt{\sigma^2 / (X_1^\top M_2 X_1)}, \text{ or} \\ \beta_{1,0} &> \hat{\beta}_1 + z_{1-\alpha/2} \sqrt{\sigma^2 / (X_1^\top M_2 X_1)}. \end{aligned}$$

Equivalently, we reject if

$$\left| \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\sigma^2 / (X_1^\top M_2 X_1)}} \right| > z_{1-\alpha/2}. \tag{4}$$

Such a test is called *two-sided* since, under the alternative, the true value of β_1 may be smaller or larger than $\beta_{1,0}$.

The expression on the left-hand side is the test statistic. To compute the rejection probability, assume that the true value is β_1 and write

$$\frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\sigma^2 / (X_1^\top M_2 X_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2 / (X_1^\top M_2 X_1)}} + \frac{\beta_1 - \beta_{1,0}}{\sqrt{\sigma^2 / (X_1^\top M_2 X_1)}}. \tag{5}$$

Then

$$\frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\sigma^2 / (X_1^\top M_2 X_1)}} \mid X \sim N \left(\frac{\beta_1 - \beta_{1,0}}{\sqrt{\sigma^2 / (X_1^\top M_2 X_1)}}, 1 \right).$$

If the null hypothesis is true, then $\beta_1 - \beta_{1,0} = 0$, and the test statistic has a standard normal distribution. In this case, by the definition of $z_{1-\alpha/2}$,

$$\begin{aligned} \Pr(\text{Reject } H_0 \mid X, H_0 \text{ true}) &= \Pr \left(\left| \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2 / (X_1^\top M_2 X_1)}} \right| > z_{1-\alpha/2} \mid X \right) \\ &= \alpha. \end{aligned}$$

Thus, the suggested test has the correct size α . If the null hypothesis is false, the distribution of the test statistic is not centered around zero, and we will see rejection rates higher than α .

The probability of rejection is a function of the true value β_1 and depends on the magnitude of the second term in (5), $|\beta_1 - \beta_{1,0}| / \sqrt{\sigma^2 / (X_1^\top M_2 X_1)}$. For example, suppose that

$$\begin{aligned}\beta_{1,0} &= 0, \\ \sqrt{\sigma^2 / (X_1^\top M_2 X_1)} &= 1, \\ \alpha &= 0.05 \quad (z_{1-\alpha/2} = 1.96).\end{aligned}$$

Let $Z \sim N(0, 1)$. In this case, the *power function* of the test is

$$\begin{aligned}\pi(\beta_1) &= \Pr\left(\left|\frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\sigma^2 / (X_1^\top M_2 X_1)}}\right| > z_{1-\alpha/2} \mid X\right) \\ &= \Pr\left(\left|\frac{\hat{\beta}_1 - \beta_1 + \beta_1 - \beta_{1,0}}{\sqrt{\sigma^2 / (X_1^\top M_2 X_1)}}\right| > 1.96 \mid X\right) \\ &= \Pr(|Z + \beta_1| > 1.96) \\ &= \Pr(Z < -1.96 - \beta_1) + \Pr(Z > 1.96 - \beta_1).\end{aligned}$$

For example,

$$\pi(\beta_1) = \begin{cases} 0.52 & \text{for } \beta_1 = -2, \\ 0.17 & \text{for } \beta_1 = -1, \\ 0.05 & \text{for } \beta_1 = 0, \\ 0.17 & \text{for } \beta_1 = 1, \\ 0.52 & \text{for } \beta_1 = 2. \end{cases}$$

In this case, the power function is minimized at $\beta_1 = \beta_{1,0}$, where $\pi(\beta_1) = \alpha$.

These power calculations assume the significance level is known in advance. A closely related question is how to summarize the strength of evidence against the null for a given realization of the test statistic. This is the role of the *p*-value. To illustrate, suppose that, given the data, the test statistic in (4) is equal to 1.88. For the standard normal distribution, $\Pr(Z > 1.88) = 0.03$. Therefore, the *p*-value for the two-sided test is $2 \times 0.03 = 0.06$. The null is rejected at any significance level exceeding 0.06.

In the case of unknown σ^2 , one can test (3) by considering the *t*-statistic:

$$\begin{aligned}T &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{s^2 / (X_1^\top M_2 X_1)}} \\ &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1 \mid X)}}.\end{aligned}\tag{6}$$

The test is given by the following decision rule:

$$\text{Reject } H_0 \text{ if } |T| > t_{n-k, 1-\alpha/2}.$$

The *t*-distribution result for the ratio of the OLS estimator to its standard error (Lecture 4) then implies that, under H_0 , $\Pr(|T| > t_{n-k, 1-\alpha/2} \mid X, H_0 \text{ true}) = \alpha$.

One can also consider *one-sided* tests. In the case of one-sided tests, the null and alternative hypotheses may be specified as

$$\begin{aligned}H_0 &: \beta_1 \leq \beta_{1,0}, \\ H_1 &: \beta_1 > \beta_{1,0}.\end{aligned}$$

Both H_0 and H_1 are composite, and the probability of rejection varies not only across the values of β_1 specified under H_1 but also across H_0 . In this case, a valid test should satisfy the following condition:

$$\sup_{\beta_1 \leq \beta_{1,0}} \Pr(\text{reject } H_0 \mid X, \beta_1) \leq \alpha, \quad (7)$$

that is, the maximum probability of rejecting H_0 when it is true should not exceed α . Let T be as defined in (6) and consider the following test (decision rule):

$$\text{Reject } H_0 \text{ when } T > t_{n-k,1-\alpha}.$$

Under H_0 , we have:

$$\begin{aligned} \Pr(\text{reject } H_0 \mid X, \beta_1 \leq \beta_{1,0}) &= \Pr(T > t_{n-k,1-\alpha} \mid X, \beta_1 \leq \beta_{1,0}) \\ &= \Pr\left(\frac{\widehat{\beta}_1 - \beta_{1,0}}{\sqrt{s^2 / (X_1^\top M_2 X_1)}} > t_{n-k,1-\alpha} \mid X, \beta_1 \leq \beta_{1,0}\right) \\ &\leq \Pr\left(\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{s^2 / (X_1^\top M_2 X_1)}} > t_{n-k,1-\alpha} \mid X, \beta_1 \leq \beta_{1,0}\right) \quad (\text{since } \beta_1 \leq \beta_{1,0}) \\ &= \alpha \quad (\text{since } \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{s^2 / (X_1^\top M_2 X_1)}} \mid X \sim t_{n-k}). \end{aligned}$$

Thus, the size control condition (7) is satisfied. Since this is a one-sided test, the entire rejection probability is assigned to the right tail of the distribution.

Testing a single linear restriction

Consider the normal linear regression model defined by Assumptions (A1)–(A5):

$$Y = X\beta + U.$$

Suppose we want to test

$$\begin{aligned} H_0 &: c^\top \beta = r, \\ H_1 &: c^\top \beta \neq r. \end{aligned}$$

In this case, c is a k -vector, r is a scalar, and under the null hypothesis

$$c_1\beta_1 + \dots + c_k\beta_k - r = 0.$$

For example, by setting $c_1 = 1$, $c_2 = -1$, $c_3 = \dots = c_k = 0$, and $r = 0$, one can test the hypothesis that $\beta_1 = \beta_2$.

The OLS estimator of β satisfies

$$\widehat{\beta} \mid X \sim N\left(\beta, \sigma^2 (X^\top X)^{-1}\right). \quad (8)$$

Then,

$$\frac{c^\top \widehat{\beta} - c^\top \beta}{\sqrt{\sigma^2 c^\top (X^\top X)^{-1} c}} \mid X \sim N(0, 1).$$

Therefore, under H_0 ,

$$\frac{c^\top \widehat{\beta} - r}{\sqrt{\sigma^2 c^\top (X^\top X)^{-1} c}} \mid X \sim N(0, 1). \quad (9)$$

Consider the t -statistic

$$\begin{aligned} T &= \frac{c^\top \widehat{\beta} - r}{\sqrt{s^2 c^\top (X^\top X)^{-1} c}} \\ &= \left(\frac{c^\top \widehat{\beta} - r}{\sqrt{\sigma^2 c^\top (X^\top X)^{-1} c}} \right) / \sqrt{\frac{U^\top M_X U}{\sigma^2} / (n - k)}. \end{aligned}$$

Under H_0 , the result in (9) holds. Further, conditional on X ,

$$U^\top M_X U / \sigma^2 \mid X \sim \chi_{n-k}^2 \text{ and independent of } \widehat{\beta}. \quad (10)$$

Therefore, under H_0 ,

$$T \mid X \sim t_{n-k}.$$

Thus, the significance-level- α two-sided test of $H_0 : c^\top \beta = r$ is given by

$$\text{Reject } H_0 \text{ if } |T| > t_{n-k, 1-\alpha/2}.$$

By setting the j -th element of c to $c_j = 1$ and the rest of the elements of c equal to zero, one obtains the test discussed in the previous section:

$$\begin{aligned} H_0 &: \beta_j = r, \\ H_1 &: \beta_j \neq r. \end{aligned}$$

One rejects H_0 if

$$\begin{aligned} |T| &= \left| \frac{\widehat{\beta}_j - r}{\sqrt{s^2 [(X^\top X)^{-1}]_{jj}}} \right| \\ &> t_{n-k, 1-\alpha/2}, \end{aligned}$$

where $[(X^\top X)^{-1}]_{jj}$ denotes the element (j, j) of the matrix $(X^\top X)^{-1}$.

Testing multiple linear restrictions

Suppose we want to test

$$\begin{aligned} H_0 &: R\beta = r, \\ H_1 &: R\beta \neq r, \end{aligned}$$

where R is a $q \times k$ matrix and r is a q -vector. For example,

- $R = I_k$, $r = 0$. In this case, we test that $\beta_1 = \dots = \beta_k = 0$.
- $R = \begin{pmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \end{pmatrix}$, $r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. In this case, $H_0 : \beta_1 + \beta_2 = 1, \beta_3 = 0$.

Consider the F -statistic

$$F = \left(R\hat{\beta} - r \right)^\top \left(s^2 R (X^\top X)^{-1} R^\top \right)^{-1} \left(R\hat{\beta} - r \right) / q.$$

We show next that under H_0 ,

$$F | X \sim F_{q, n-k}. \quad (11)$$

First, it follows from (8) that

$$R\hat{\beta} | X \sim N \left(R\beta, \sigma^2 R (X^\top X)^{-1} R^\top \right).$$

Then, under H_0 ,

$$R\hat{\beta} - r | X \sim N \left(0, \sigma^2 R (X^\top X)^{-1} R^\top \right).$$

Further, by Lemma 2 in Lecture 4,

$$\left(R\hat{\beta} - r \right)^\top \left(\sigma^2 R (X^\top X)^{-1} R^\top \right)^{-1} \left(R\hat{\beta} - r \right) | X \sim \chi_q^2.$$

The result in (11) follows from (10) and the definition of the F -distribution. The test is given by

$$\begin{aligned} \text{Reject } H_0 \text{ if } F &= \left(R\hat{\beta} - r \right)^\top \left(s^2 R (X^\top X)^{-1} R^\top \right)^{-1} \left(R\hat{\beta} - r \right) / q \\ &> F_{q, n-k, 1-\alpha}. \end{aligned}$$

Restricted OLS

An alternative approach to hypothesis testing is based on restricted estimation. One might consider the loss of fit resulting from choosing values other than $\hat{\beta}$ for the regression coefficients. Consider the *restricted* LS problem

$$\min_b (Y - Xb)^\top (Y - Xb) \quad \text{s.t.} \quad Rb = r.$$

A Lagrangian function for this problem is

$$L(b, \lambda) = (Y - Xb)^\top (Y - Xb) + 2\lambda^\top (Rb - r),$$

where λ is a q -vector. Let $\tilde{\beta}, \tilde{\lambda}$ be the solution, where $\tilde{\beta}$ is the restricted LS estimator. It has to satisfy the first-order conditions

$$\frac{\partial L(\tilde{\beta}, \tilde{\lambda})}{\partial b} = 2X^\top X\tilde{\beta} - 2X^\top Y + 2R^\top \tilde{\lambda} = 0, \quad (12)$$

$$\frac{\partial L(\tilde{\beta}, \tilde{\lambda})}{\partial \lambda} = R\tilde{\beta} - r = 0. \quad (13)$$

From (12),

$$\begin{aligned} \tilde{\beta} &= (X^\top X)^{-1} \left(X^\top Y - R^\top \tilde{\lambda} \right) \\ &= \hat{\beta} - (X^\top X)^{-1} R^\top \tilde{\lambda}. \end{aligned}$$

Combining the last equation with (13),

$$\begin{aligned} r &= R\tilde{\beta} \\ &= R\hat{\beta} - R(X^\top X)^{-1} R^\top \tilde{\lambda}, \end{aligned}$$

and

$$\tilde{\lambda} = \left(R (X^\top X)^{-1} R^\top \right)^{-1} (R\hat{\beta} - r).$$

Therefore, the restricted LS estimator is given by

$$\tilde{\beta} = \hat{\beta} - (X^\top X)^{-1} R^\top \left(R (X^\top X)^{-1} R^\top \right)^{-1} (R\hat{\beta} - r).$$

Define the restricted residuals

$$\begin{aligned} \tilde{U} &= Y - X\tilde{\beta} \\ &= (Y - X\hat{\beta}) + X (X^\top X)^{-1} R^\top \left(R (X^\top X)^{-1} R^\top \right)^{-1} (R\hat{\beta} - r) \\ &= \hat{U} + X (X^\top X)^{-1} R^\top \left(R (X^\top X)^{-1} R^\top \right)^{-1} (R\hat{\beta} - r), \end{aligned}$$

where \hat{U} is the vector of unrestricted residuals. Consider the *restricted* Residual Sum of Squares:

$$\begin{aligned} RSS_r &= \tilde{U}^\top \tilde{U} \\ &= \hat{U}^\top \hat{U} + (R\hat{\beta} - r)^\top \left(R (X^\top X)^{-1} R^\top \right)^{-1} (R\hat{\beta} - r) \\ &\quad + 2\hat{U}^\top X (X^\top X)^{-1} R^\top \left(R (X^\top X)^{-1} R^\top \right)^{-1} (R\hat{\beta} - r) \\ &= RSS + (R\hat{\beta} - r)^\top \left(R (X^\top X)^{-1} R^\top \right)^{-1} (R\hat{\beta} - r), \end{aligned}$$

where $RSS = \hat{U}^\top \hat{U}$ denotes the unrestricted Residual Sum of Squares. Since $s^2 = \hat{U}^\top \hat{U} / (n - k)$, the F -statistic discussed in the previous section can be written as

$$F = \frac{(RSS_r - RSS) / q}{RSS / (n - k)}. \quad (14)$$

Examples:

1. Model significance. Consider a model with an intercept

$$Y_i = \beta_1 + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + U_i.$$

Consider the null hypothesis $H_0 : \beta_2 = \dots = \beta_k = 0$. The restricted model is given by

$$Y_i = \beta_1 + U_i.$$

In this case, the restricted LS estimator is $\tilde{\beta}_1 = n^{-1} \sum_{i=1}^n Y_i = \bar{Y}$, and $RSS_r = TSS = \sum_{i=1}^n (Y_i - \bar{Y})^2$. Therefore,

$$\begin{aligned} F &= \frac{(TSS - RSS) / (k - 1)}{RSS / (n - k)} \\ &= \frac{ESS / (k - 1)}{RSS / (n - k)} \\ &= \frac{R^2 / (k - 1)}{(1 - R^2) / (n - k)} \\ &\sim F_{k-1, n-k} \quad \text{under } H_0. \end{aligned}$$

2. Consider the model

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + U_i,$$

and the null hypothesis $H_0 : \beta_2 = \beta_3$. The restricted model is given by

$$Y_i = \beta_1 + \beta_2 (X_{i2} + X_{i3}) + U_i.$$

Thus, to test whether $\beta_2 = \beta_3$, one may construct the new variable $W_i = (X_{i2} + X_{i3})$, compute RSS_r by taking the RSS from the regression of Y_i on a constant and W_i , compute RSS from the unrestricted regression, and construct the F -statistic according to (14).