December 9, 2019

Hypothesis Testing

Basic concepts

Let $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ be a parameter of interest. Some examples of $\boldsymbol{\theta}$ include:

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

A statistical hypothesis is an assertion about $\boldsymbol{\theta}$. 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 $\Theta_0 \cap \Theta_1 = \varnothing$ and $\Theta_0 \cup \Theta_1 = \Theta$. The two competing hypotheses are:

- Null hypothesis $H_0 : \boldsymbol{\theta} \in \Theta_0$. This is a hypothesis that is held as true, unless data provides *sufficient* evidence against it.
- Alternative hypothesis $H_1 : \boldsymbol{\theta} \in \Theta_1$. 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" and the case that he is interested in is stated as H_1.

Note that the two hypotheses, H_0 and H_1 must be *disjoint*. Their union defines the *maintained* hypothesis, i.e. the space of values that $\boldsymbol{\theta}$ 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 $H_0 : \boldsymbol{\theta} \in \Theta_0$ is a *simple* hypothesis. Otherwise, we say that H_0 is a composite hypothesis. Similarly, $H_1 : \boldsymbol{\theta} \in \Theta_1$ 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 *accept* or *reject* H_0 is based on a *test statistic*, which is a function of data ($\boldsymbol{X}$ and $\boldsymbol{Y}$ in the case of a regression model). Let $S \in \mathcal{S}$ denote a statistic and the range of its values. A decision rule is defined by a partition of $\mathcal{S}$ into acceptance region $\mathcal{A}$ and rejection (critical) region $\mathcal{R}$. Note that the acceptance and rejection regions must be disjoint ($\mathcal{A} \cap \mathcal{R} = \varnothing$), 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 accepting 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:

$$\pi(\boldsymbol{\theta}) = \Pr_{\boldsymbol{\theta}}(S \in \mathcal{R}),$$

where $\Pr_{\boldsymbol{\theta}}(\cdot)$ denotes that the probability must be calculated under the assumption that the true value of the parameter is $\boldsymbol{\theta}$. Thus, a power function of a test gives the probability of rejecting H_0 for every possible value of $\boldsymbol{\theta}$. The largest probability of Type I error (rejecting H_0 when it is true) is

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

The expression above is also called the *size* of a test. When H_0 is simple, i.e. $\Theta = \{\boldsymbol{\theta}_0\}$, the size can be computed simply as $\pi(\boldsymbol{\theta}_0) = \Pr_{\boldsymbol{\theta}_0}(S \in \mathcal{R})$.

The probability of Type II error (accepting H_0 when it is false) is:

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

Typically, Θ_1 has many elements, and therefore the probability of Type II error depends on the true value $\boldsymbol{\theta}$. One would like to have the probabilities of Type I and 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 his preferred hypothesis as H_1).

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

Note that size α tests are level α tests. 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 significance levels. 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 a test is called the p-value. Instead of reporting test outcomes (accept or 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 *greater* than the p-value.

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

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

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

Finite Sample Theory

We discuss the finite sample tests for the normal classical linear regression defined by

(A1) $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$.

(A2) $\mathbb{E}\left(\boldsymbol{e}|\boldsymbol{X}\right) = 0$

(A3) $\mathrm{Var}\left(\boldsymbol{e}|\boldsymbol{X}\right) = \sigma^2 \boldsymbol{I}_n$ ($\sigma^2 > 0$ is a constant.)

(A4) $\mathrm{rank}(\boldsymbol{X}) = k$.

(A5) $\boldsymbol{e}|\boldsymbol{X} \sim N\left(0, \sigma^2 \boldsymbol{I}\right)$.

Testing a hypothesis about a single coefficient

Consider the partitioned regression

$$\boldsymbol{Y} = \beta_1 \boldsymbol{X}_1 + \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{e},$$

where $\boldsymbol{X}_1$ is the $n \times 1$ vector of the observations of the first regressor. Assume that the variance of the error term $\sigma^2 = \mathbb{E}e_i^2$ is known. Let $\widehat{\beta}_1$ 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 a α-level test can be based on the $CI_{1-\alpha}$. The $1-\alpha$ level confidence interval for β_1 is

$$CI_{1-\alpha} = \left[\widehat{\beta}_1 - z_{1-\alpha/2}\sqrt{\sigma^2 / \left(\boldsymbol{X}_1' \boldsymbol{M}_2 \boldsymbol{X}_1\right)}, \widehat{\beta}_1 + z_{1-\alpha/2}\sqrt{\sigma^2 / \left(\boldsymbol{X}_1' \boldsymbol{M}_2 \boldsymbol{X}_1\right)}\right].$$

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

$$\begin{aligned} \beta_{1,0} &< \widehat{\beta}_1 - z_{1-\alpha/2}\sqrt{\sigma^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)}, \text{ or} \\ \beta_{1,0} &> \widehat{\beta}_1 + z_{1-\alpha/2}\sqrt{\sigma^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)}. \end{aligned}$$

Equivalently, we reject if

$$\left|\frac{\widehat{\beta}_1 - \beta_{1,0}}{\sqrt{\sigma^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)}}\right| > z_{1-a/2}. \tag{4}$$

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 a test statistic. In order to compute the probability to reject the null, let's assume that the true value is given by β_1. Write

$$\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)}} + \frac{\beta_1 - \beta_{1,0}}{\sqrt{\sigma^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)}}. \tag{5}$$

We have that

$$Z = \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)}}|\boldsymbol{X} \sim N\left(0,1\right)$$

and

$$\frac{\widehat{\beta}_1 - \beta_{1,0}}{\sqrt{\sigma^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)}}|\boldsymbol{X} \sim N\left(\frac{\beta_1 - \beta_{1,0}}{\sqrt{\sigma^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)}}, 1\right).$$

Note that Z is independent from $\boldsymbol{X}$ and thus $\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1$. 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 $z_{1-a/2}$,

$$\begin{aligned} \Pr\left(\text{Reject } H_0|H_0 \text{ is true}\right) &= \Pr\left(\left|\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)}}\right| > z_{1-a/2}\right) \\ &= \alpha. \end{aligned}$$

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 to reject is a function of the true value β_1 and depends on the magnitude of the second term in (5), $\left|\beta_1 - \beta_{1,0}\right| / \sqrt{\sigma^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)}$. Now

$$\begin{aligned} \pi\left(\beta_1\right) &= \Pr\left(\left|\frac{\widehat{\beta}_1 - \beta_{1,0}}{\sqrt{\sigma^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)}}\right| > z_{1-a/2}\right) \\ &= \Pr\left(\left|\frac{\widehat{\beta}_1 - \beta_1 + \beta_1 - \beta_{1,0}}{\sqrt{\sigma^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)}}\right| > z_{1-a/2}\right) \\ &= \Pr\left(\left|Z + \frac{\beta_1 - \beta_{1,0}}{\sqrt{\sigma^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)}}\right| > z_{1-a/2}\right). \end{aligned}$$

For example, suppose that

$$\begin{aligned}\beta_{1,0} &= 0,\\ \sqrt{\sigma^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)} &= 1,\\ \alpha &= 0.05,\ (\text{and } z_{1-\alpha/2}=1.96).\end{aligned}$$

In this case, the *power function* of the test is

$$\begin{aligned}\pi\left(\beta_1\right) &= \Pr\left(\left|\frac{\widehat{\beta}_1-\beta_{1,0}}{\sqrt{\sigma^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)}}\right|>z_{1-a/2}\right)\\ &= \Pr\left(\left|Z+\beta_1\right|>1.96\right)\\ &= \Pr\left(Z<-1.96-\beta_1\right)+P\left(Z>1.96-\beta_1\right).\end{aligned}$$

For example,

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

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

For p-value calculation, consider the following example. Suppose that given the data, the test statistic in (4) is equal 1.88. For the standard normal distribution, $P\left(Z>1.88\right)=0.03$. Therefore, the p-value for a *two-sided* test is 0.06. One would reject the null for all tests with significance level higher than 0.06.

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

$$\begin{aligned}T &= \frac{\widehat{\beta}_1-\beta_{1,0}}{\sqrt{s^2/\left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)}}\\ &= \frac{\widehat{\beta}_1-\beta_{1,0}}{\sqrt{\widehat{\mathrm{Var}}\left(\widehat{\beta}_1|\boldsymbol{X}\right)}}.\end{aligned} \tag{6}$$

The test is given by the following decision rule:

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

We have shown that if H_0 is true, $T|\boldsymbol{X}\sim t_{n-k}$ and thus T is independent from $\boldsymbol{X}$. Under H_0, $\Pr\left(|T|>t_{n-k,1-\alpha/2}|H_0 \text{ is true}\right)=\alpha$.

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

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

Note that in this case, 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:

$$\sup_{\beta_1 \leq \beta_{1,0}} \Pr\left(\text{reject } H_0 | \beta_1\right) \leq \alpha, \tag{7}$$

i.e. the maximum probability to reject 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:

$$\begin{aligned}
\Pr\left(\text{reject } H_0 | \beta_1 \leq \beta_{1,0}\right) &= \Pr\left(T > t_{n-k,1-\alpha} | X, \beta_1 \leq \beta_{1,0}\right) \\
&= \Pr\left(\frac{\widehat{\beta}_1 - \beta_{1,0}}{\sqrt{s^2 / \left(\boldsymbol{X}_1' \boldsymbol{M}_2 \boldsymbol{X}_1\right)}} > t_{n-k,1-\alpha} | \beta_1 \leq \beta_{1,0}\right) \\
&\leq \Pr\left(\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{s^2 / \left(\boldsymbol{X}_1' \boldsymbol{M}_2 \boldsymbol{X}_1\right)}} > t_{n-k,1-\alpha} | \beta_1 \leq \beta_{1,0}\right) \text{ (since } \beta_1 \leq \beta_{1,0}\text{)} \\
&= \alpha \text{ (since } \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{s^2 / \left(\boldsymbol{X}_1' \boldsymbol{M}_2 \boldsymbol{X}_1\right)}} \sim t_{n-k}\text{)}.
\end{aligned}$$

Thus, the size control condition (7) is satisfied. Note, since this is a one-sided test, the probability of type I error is assigned only to the right tail of the distribution.

Testing a single linear restriction

Suppose we want to test

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

In this case, $\boldsymbol{c}$ is a k-vector, r is a scalar, and under the null hypothesis

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

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

We have that the LS estimator of $\boldsymbol{\beta}$

$$\widehat{\boldsymbol{\beta}} | \boldsymbol{X} \sim N\left(\boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right). \tag{8}$$

Then,

$$\frac{\boldsymbol{c}'\widehat{\boldsymbol{\beta}} - \boldsymbol{c}'\boldsymbol{\beta}}{\sqrt{\sigma^2 \boldsymbol{c}' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{c}}} | \boldsymbol{X} \sim N\left(0, 1\right).$$

Therefore, under H_0,

$$\frac{\boldsymbol{c}'\widehat{\boldsymbol{\beta}} - r}{\sqrt{\sigma^2 \boldsymbol{c}' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{c}}} | \boldsymbol{X} \sim N(0,1). \tag{9}$$

Consider the t-statistic

$$\begin{aligned} T &= \frac{\boldsymbol{c}'\widehat{\boldsymbol{\beta}} - r}{\sqrt{s^2 \boldsymbol{c}' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{c}}} \\ &= \left(\frac{c'\widehat{\beta} - r}{\sqrt{\sigma^2 \boldsymbol{c}' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{c}}} \right) / \sqrt{\frac{\boldsymbol{e}'\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{e}}{\sigma^2} / (n-k)}. \end{aligned}$$

Under H_0, the result in (9) holds. Further, conditional on $\boldsymbol{X}$,

$$\boldsymbol{e}'\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{e}/\sigma^2 | \boldsymbol{X} \sim \chi^2_{n-k} \text{ and independent of } \widehat{\boldsymbol{\beta}}. \tag{10}$$

Therefore, under H_0,

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

Thus, the significance level α two-sided test of $H_0 : \boldsymbol{c}'\boldsymbol{\beta} = r$ is given by

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

By setting the j-th element of $\boldsymbol{c}$, $c_j = 1$ and the rest of the elements of equal $\boldsymbol{c}$ 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

$$\begin{aligned} |T| &= \left| \frac{\widehat{\beta}_j - r}{\sqrt{s^2 \left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \right]_{jj}}} \right| \\ &> t_{n-k,1-\alpha/2}, \end{aligned}$$

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

Testing multiple linear restrictions

Suppose we want to test

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

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

- $\boldsymbol{R} = \boldsymbol{I}_k$, $\boldsymbol{r} = \boldsymbol{0}$. In this case, we test that $\beta_1 = \ldots = \beta_k = 0$.

- $R = \begin{pmatrix} 1 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \end{pmatrix}$, $r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. In this case, $H_0 : \beta_1 + \beta_2 = 1, \beta_3 = 0$.

Consider the F-statistic

$$F = \frac{(RSS_r - RSS_{ur})/q}{RSS_{ur}/(n-k)} = \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)' \left(s^2 \boldsymbol{R} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{R}'\right)^{-1} \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)/q.$$

We show next that under H_0,

$$F|\boldsymbol{X} \sim F_{q,n-k}. \tag{11}$$

First, it follows from (8),

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

Then, under H_0,

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

It follows that

$$\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)' \left(\sigma^2 \boldsymbol{R} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{R}'\right)^{-1} \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)|\boldsymbol{X} \sim \chi^2_q.$$

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

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

Consider a model with the intercept

$$Y_i = \beta_1 + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + U_i,$$

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

$$Y_i = \beta_1 + U_i.$$

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

$$\begin{aligned} F &= \frac{(TSS - RSS_{ur})/(k-1)}{RSS_{ur}/(n-k)} \\ &= \frac{ESS/(k-1)}{RSS_{ur}/(n-k)} \\ &= \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \\ &\sim F_{k-1,n-k}. \end{aligned}$$

Large Sample Theory

We discuss large sample properties for the linear regression model defined by the following assumptions:

(A1) $Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + e_i$, for all $i = 1, ..., n$.

(A2) $\mathbb{E}(e_i \boldsymbol{X}_i) = \boldsymbol{0}$.

(A3) $\{(Y_i, \boldsymbol{X}_i) : i = 1, \ldots, n\}$ are iid.

(A4) $\mathbb{E}(\boldsymbol{X}_i \boldsymbol{X}_i')$ is a finite positive definite matrix.

(A5) $\mathbb{E}X_{i,j}^4 < \infty$ for all $j = 1, \ldots, k$.

(A6) $\mathbb{E}e_i^4 < \infty$.

(A7) $\mathbb{E}\left(e_i^2 \boldsymbol{X}_i \boldsymbol{X}_i'\right)$ is positive definite.

In this section we discuss asymptotic tests of the null hypothesis $H_0 : \boldsymbol{h}(\boldsymbol{\beta}) = \boldsymbol{0}$ against the alternative $H_1 : \boldsymbol{h}(\boldsymbol{\beta}) \neq \boldsymbol{0}$, where $\boldsymbol{h} : \mathbb{R}^k \to \mathbb{R}^q$ is a continuously differentiable function in the neighborhood of $\boldsymbol{\beta}$.

The restriction under H_0 includes the linear restrictions discussed as a special case (set $\boldsymbol{h}(\boldsymbol{\beta}) = \boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{r}$). We consider a *Wald test statistic*:

$$\begin{aligned} W_n &= n\boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n\right)' \left(\widehat{\text{AsyVar}}\left(\boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n\right)\right)\right)^{-1} \boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n\right) \\ &= n\boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n\right)' \left(\frac{\partial \boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n\right)}{\partial \boldsymbol{\beta}'} \widehat{\boldsymbol{V}}_n \frac{\partial \boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n\right)'}{\partial \boldsymbol{\beta}}\right)^{-1} \boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n\right), \end{aligned}$$

where AsyVar denotes the asymptotic variance. The asymptotic α-size test of $H_0 : \boldsymbol{h}(\boldsymbol{\beta}) = \boldsymbol{0}$ is given by

$$\text{Reject } H_0 \text{ if } W_n > \chi^2_{q,1-\alpha},$$

where $\chi^2_{q,1-\alpha}$ is the $(1-\alpha)$ quantile of the χ^2_q distribution. A test based on W_n is called *consistent* if $\Pr\left(W_n > \chi^2_{q,1-\alpha} | H_1 \text{ is true}\right) \to 1$.

Theorem 1. Under Assumptions (A1) - (A7),

(a) $\Pr\left(W_n > \chi^2_{q,1-\alpha} | H_0 \text{ is true}\right) \to \alpha$.

(b) $\Pr\left(W_n > \chi^2_{q,1-\alpha} | H_1 \text{ is true}\right) \to 1$.

Proof. (a) Since $n^{1/2}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) \to_d N(\boldsymbol{0}, \boldsymbol{V}_{\boldsymbol{\beta}})$ and $\boldsymbol{h}$ is continuous at $\boldsymbol{\beta}$ under H_0, and by the delta method,

$$n^{1/2}\boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n\right) \to_d N\left(\boldsymbol{0}, \frac{\partial \boldsymbol{h}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \boldsymbol{V}_{\boldsymbol{\beta}} \frac{\partial \boldsymbol{h}(\boldsymbol{\beta})'}{\partial \boldsymbol{\beta}}\right).$$

Furthermore, we have that

$$\begin{aligned} \frac{\partial \boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n\right)}{\partial \boldsymbol{\beta}'} &\to_p \frac{\partial \boldsymbol{h}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}, \\ \widehat{\boldsymbol{V}}_{\boldsymbol{\beta}} &\to_p \boldsymbol{V}_{\boldsymbol{\beta}}. \end{aligned}$$

By Slutsky's Theorem, under H_0,

$$\begin{aligned}\left(\frac{\partial \boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n\right)}{\partial \boldsymbol{\beta}'}\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}\frac{\partial \boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n\right)'}{\partial \boldsymbol{\beta}}\right)^{-1/2} n^{1/2}\boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n\right) &\to_d \left(\frac{\partial \boldsymbol{h}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\boldsymbol{V}_{\boldsymbol{\beta}}\frac{\partial \boldsymbol{h}(\boldsymbol{\beta})'}{\partial \boldsymbol{\beta}}\right)^{-1/2} N\left(0, \frac{\partial \boldsymbol{h}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\boldsymbol{V}_{\boldsymbol{\beta}}\frac{\partial \boldsymbol{h}(\boldsymbol{\beta})'}{\partial \boldsymbol{\beta}}\right) \\ &= N\left(\mathbf{0}, \boldsymbol{I}_q\right).\end{aligned}$$

Then, by the Continuous Mapping Theorem, under H_0,

$$W_n \to_d \chi^2_q,$$

which completes the proof of part (a).

(b) Under the alternative, $\boldsymbol{h}(\boldsymbol{\beta}) \neq \mathbf{0}$. Hence, by the Continuous Mapping Theorem,

$$\begin{aligned}\boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n\right) &\to_p \boldsymbol{h}(\boldsymbol{\beta}) \\ &\neq \mathbf{0}.\end{aligned}$$

Therefore,

$$W_n/n \to_p \boldsymbol{h}(\boldsymbol{\beta})'\left(\frac{\partial \boldsymbol{h}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\boldsymbol{V}_{\boldsymbol{\beta}}\frac{\partial \boldsymbol{h}(\boldsymbol{\beta})'}{\partial \boldsymbol{\beta}}\right)^{-1}\boldsymbol{h}(\boldsymbol{\beta}) > 0.$$

Set $\epsilon = \boldsymbol{h}(\boldsymbol{\beta})'\left(\frac{\partial \boldsymbol{h}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\boldsymbol{V}_{\boldsymbol{\beta}}\frac{\partial \boldsymbol{h}(\boldsymbol{\beta})'}{\partial \boldsymbol{\beta}}\right)^{-1}\boldsymbol{h}(\boldsymbol{\beta})/2$. Now

$$\begin{aligned}\Pr\left(\frac{W_n}{n} > \frac{\chi^2_{q,1-\alpha}}{n}\right) &\geq \Pr\left(\frac{W_n}{n} > \epsilon\right) \\ &\geq \Pr\left(\left|\frac{W_n}{n} - \boldsymbol{h}(\boldsymbol{\beta})'\left(\frac{\partial \boldsymbol{h}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\boldsymbol{V}_{\boldsymbol{\beta}}\frac{\partial \boldsymbol{h}(\boldsymbol{\beta})'}{\partial \boldsymbol{\beta}}\right)^{-1}\boldsymbol{h}(\boldsymbol{\beta})\right| < \epsilon\right) \\ &\to 1,\end{aligned}$$

where the first inequality holds whenever n is sufficiently large.

Note that in the case of a linear restriction $\boldsymbol{h}(\boldsymbol{\beta}) = \boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{r}$, we have that:

$$W_n = n\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}_n - \boldsymbol{r}\right)'\left(\boldsymbol{R}\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}\boldsymbol{R}'\right)^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}_n - \boldsymbol{r}\right).$$

Further, in the homoskedastic case, one can replace $\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}$ by $s^2\left(\boldsymbol{X}'\boldsymbol{X}/n\right)^{-1}$. Then, the Wald statistic becomes

$$W_n = \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}_n - \boldsymbol{r}\right)'\left(s^2\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right)^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}_n - \boldsymbol{r}\right),$$

which is a similar expression to that of the F statistic, except for adjustment to the number of degrees of freedom q in the numerator.