Hypothesis testing and interval estimation in finite and large samples

Econometrics (35B206), Lecture 3

Tunga Kantarcı, TiSEM, Tilburg University, Spring 2019

Hypothesis testing in finite and large samples

Consider the LRM

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

So far we have been interested in estimating the true $\boldsymbol{\beta}$. Here we are interested in forming a hypothesis about $\boldsymbol{\beta}$, and testing it.

Hypothesis testing in finite and large samples

Let us hypothesise that the true β_k is equal to the constant β_k^0. Then, the null and the alternative hypotheses are

$$H_0 : \beta_k = \beta_k^0$$
$$H_1 : \beta_k \neq \beta_k^0$$

Hypothesis testing in finite and large samples

The hypothesis we are testing is

$$H_0 : \beta_k = \beta_k^0$$
$$H_1 : \beta_k \neq \beta_k^0.$$

That is, we want to check whether

$$\beta_k = \beta_k^0.$$

But we do not observe β_k. Hence, we cannot make this check. But we can estimate β_k. Suppose that $\hat{\beta}_k$ is the OLS estimate of β_k. Now we can check whether

$$\hat{\beta}_k = \beta_k^0.$$

Suppose that this is the case. Then, our test is complete, and we conclude that H_0 is true. But this conclusion has a problem.

Hypothesis testing in finite and large samples

$\hat{\beta}_k$ is a random variable, and it has a sampling distribution. Hence, there is a probability associated with the condition

$$\hat{\beta}_k = \beta_k^0.$$

Therefore, we need to check if this holds in statistical terms. We do this check by constructing a test statistic based on this random variable.

Hypothesis testing in finite and large samples

A test statistic is a random variable, and it has a distribution. What determines the distribution of, e.g., the t test statistic? The t statistic is a function of ε. Hence, the distribution of ε determines the distribution of the t statistic.

Hypothesis testing in finite and large samples

If we assume that ε is normal, the test statistic has an exact distribution. An exact distribution means that the distribution is valid for any finite n. For example, the t test statistic has a t distribution if ε is normal. This is a distribution tabulated in the appendix of a textbook of statistics.

Hypothesis testing in finite and large samples

If ε is not normal, the test statistic does not have an exact distribution. But if we require that n is large, then the test statistic has an asymptotic distribution that approximates an exact distribution.

Hypothesis testing in finite samples: single restriction

Recall that

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}.$$

We know from the previous lecture that if

$$\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2 \boldsymbol{I}\right]$$

then

$$\hat{\boldsymbol{\beta}} \mid \boldsymbol{X} \sim N\left[\boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right].$$

This means that

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right].$$

Hypothesis testing in finite samples: single restriction

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right].$$

Often we are interested in testing a linear restriction on a given true coefficient. E.g.,

$$H_0 : \beta_k = \beta_k^0$$
$$H_1 : \beta_k \neq \beta_k^0.$$

That is, we are interested in the element located in the row k of the $K \times 1$ vector $\boldsymbol{\beta}$, and in the observation located in the row k and column k of the $K \times K$ matrix $\boldsymbol{X}'\boldsymbol{X}$. That is, we are interested in

$$\hat{\beta}_k - \beta_k \mid \boldsymbol{X} \sim N\left[0, \sigma^2 \left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}\right].$$

Define

$$S^{kk} \equiv \left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}.$$

Hypothesis testing in finite samples: single restriction

$$\hat{\beta}_k - \beta_k \mid \boldsymbol{X} \sim N\left[0, \sigma^2 S^{kk}\right].$$

Under the null $\beta_k = \beta_k^0$. Hence,

$$\hat{\beta}_k - \beta_k^0 \mid \boldsymbol{X} \sim N\left[0, \sigma^2 S^{kk}\right],$$

assuming that $\mathrm{E}\left[\hat{\beta}_k \mid \boldsymbol{X}\right] = \beta_k^0$. We want $\hat{\beta}_k - \beta_k^0$ because we want to test whether this is equal to 0.

Hypothesis testing in finite samples: single restriction

$$\hat{\beta}_k - \beta_k^0 \mid \boldsymbol{X} \sim N\left[0, \sigma^2 S^{kk}\right].$$

Standardise $\hat{\beta}_k - \beta_k^0$ to get

$$z_k \mid \boldsymbol{X} \equiv \left. \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\sigma^2 S^{kk}}} \right| \boldsymbol{X} \sim N[0, 1].$$

The distribution does not depend on $\hat{\beta}_k$, β_k^0, σ, or $\boldsymbol{X}$. Hence,

$$z_k \sim N[0, 1].$$

This is a convenient simplification because we do not need to condition on $\boldsymbol{X}$ while using the test statistic.

Hypothesis testing in finite samples: single restriction

$$z_k = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\sigma^2 S^{kk}}}.$$

z_k is not usable because σ^2 is unknown. Replace σ^2 with its unbiased estimator

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-K}.$$

We obtain

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{\sigma}^2 S^{kk}}}.$$

But we change from z_k to t_k. So how is t_k distributed?

Hypothesis testing in finite samples: single restriction

$$
\begin{aligned}
t_k &= \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{\sigma}^2 S^{kk}}} \\
&= \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{\sigma}^2}\sqrt{S^{kk}}} \frac{\sqrt{\sigma^2}}{\sqrt{\sigma^2}} \\
&= \frac{\hat{\beta}_k - \beta_k^0 / \sqrt{\sigma^2 S^{kk}}}{\sqrt{\hat{\sigma}^2/\sigma^2}} \\
&= \frac{\left(\hat{\beta}_k - \beta_k^0\right) / \sqrt{\sigma^2 S^{kk}}}{\sqrt{\hat{\sigma}^2/\sigma^2}\sqrt{(n-K)}/\sqrt{(n-K)}} \\
&= \frac{\left(\hat{\beta}_k - \beta_k^0\right) / \sqrt{\sigma^2 S^{kk}}}{\sqrt{\left(\hat{\sigma}^2/\sigma^2\right)(n-K)/(n-K)}}.
\end{aligned}
$$

Hypothesis testing in finite samples: single restriction

$$t_k = \frac{\left(\hat{\beta}_k - \beta_k^0\right) / \sqrt{\sigma^2 S^{kk}}}{\sqrt{\left(\hat{\sigma}^2/\sigma^2\right)(n-K) / (n-K)}}.$$

As shown above, the numerator is distributed as

$$\left(\hat{\beta}_k - \beta_k^0\right) / \sqrt{\sigma^2 S^{kk}} = z_k \sim N[0,1].$$

As shown below, the denominator is distributed as

$$\left(\hat{\sigma}^2/\sigma^2\right)(n-K) \sim \chi^2[n-K].$$

As shown below, the last two terms are orthogonal (Greene, Theorem 4.6, p. 117). Then,

$$t_k \sim \frac{N[0,1]}{\left\{\chi^2[n-K] / (n-K)\right\}^{1/2}} = t[n-K].$$

Hypothesis testing in finite samples: single restriction

$$
\begin{aligned}
\frac{\hat{\sigma}^2}{\sigma^2}(n-K) &= \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-K}\frac{1}{\sigma^2}(n-K) \\
&= \frac{\hat{\varepsilon}'\hat{\varepsilon}}{\sigma^2} \\
&= \frac{(\boldsymbol{M}\varepsilon)'(\boldsymbol{M}\varepsilon)}{\sigma^2} \\
&= \frac{\varepsilon'\boldsymbol{M}'\boldsymbol{M}\varepsilon}{\sigma^2} \\
&= \frac{\varepsilon'\boldsymbol{M}\varepsilon}{\sigma^2} \\
&= \left(\frac{\varepsilon}{\sigma}\right)'\boldsymbol{M}\left(\frac{\varepsilon}{\sigma}\right).
\end{aligned}
$$

Hypothesis testing in finite samples: single restriction

Since

$$\frac{\boldsymbol{\varepsilon}}{\sigma} \sim N\left[\mathbf{0}, \boldsymbol{I}\right]$$

by Greene, Theorem B.8. we have

$$\left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right)' \boldsymbol{M} \left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right) \sim \chi^2\left[rank\left(\boldsymbol{M}\right)\right].$$

Since

$$\begin{aligned} rank\left(\boldsymbol{M}\right) &= trace\left(\boldsymbol{M}\right) \\ &= trace\left(\boldsymbol{I} - \boldsymbol{P}\right) \\ &= trace\left(\boldsymbol{I}\right) - trace\left(\boldsymbol{P}\right) \\ &= n - rank\left(\boldsymbol{X}\right) \\ &= n - K, \end{aligned}$$

we have

$$\left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right)' \boldsymbol{M} \left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right) \sim \chi^2\left[n - K\right].$$

Hypothesis testing in finite samples: single restriction

$$\frac{\varepsilon}{\sigma} \sim N\left[\mathbf{0}, \boldsymbol{I}\right].$$

The terms

$$\frac{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}}{\sigma} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\frac{\varepsilon}{\sigma}$$

and

$$\frac{\hat{\sigma}^2}{\sigma^2}\left(n-K\right) = \left(\frac{\varepsilon}{\sigma}\right)'\boldsymbol{M}\left(\frac{\varepsilon}{\sigma}\right)$$

are statistically independent if

$$\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{M} = \mathbf{0}$$

according to Greene, Theorem B.12. In fact,

$$\begin{aligned}
\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{M} &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\boldsymbol{I} - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right) \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' - \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' \\
&= \mathbf{0}.
\end{aligned}$$

Hypothesis testing in finite samples: single restriction

So the only difference between

$$z_k = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\sigma^2 S^{kk}}} \sim N[0, 1]$$

and

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{\sigma}^2 S^{kk}}} \sim t[n - K]$$

is that we have replaced the unknown σ^2 by its estimator $\hat{\sigma}^2$, and the result is that we move from a standard normal distribution to a t distribution which has slightly thicker tails.

Hypothesis testing in finite samples: single restriction

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{\sigma}^2 S^{kk}}}.$$

What is the intuition of this test statistic? Is the distance between $\hat{\beta}_k$ and β_k^0 sufficiently large, with the distance measured in terms of the sampling variance of $\hat{\beta}_k$? Is t_k sufficiently large? If it is, reject the null that $\beta_k = \beta_k^0$. This is the decision rule of the test.

Hypothesis testing in finite samples: single restriction

How large t_k should be depends on the threshold t value we want to consider. This threshold t value, or the critical t value, is

$$t^c_{\frac{\alpha}{2},n-K}.$$

t^c is a value from the t distribution and depends on the following:

$\frac{\alpha}{2}$: area under the t distribution covering up to where we want t^c to rest. Hence, $\frac{\alpha}{2}$ determines the t^c we want to consider. It is often taken as 5%. So "statisticians are people whose aim in life is to be wrong 5% of the time!" (Kempthorne and Doerfler, 1969).

$n - K$: degrees of freedom which determines the shape of the t distribution.

Hypothesis testing in finite samples: single restriction

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{\sigma}^2 S^{kk}}}.$$

One point left unclear is that the distance $\hat{\beta}_k - \beta_k^0$ can be positive or negative. Hence, t_k can be positive or negative.

If t_k is positive, we would reject the null if

$$t_k > t^c_{1-\frac{\alpha}{2},n-K}.$$

If t_k is negative, we would reject the null if

$$t_k < t^c_{\frac{\alpha}{2},n-K}.$$

Hypothesis testing in finite samples: single res. example

Mincer (1974) considers the regression of the log of wage on exper, educ, and IQ score. The data contains 935 observations. Estimation of this regression gives

$$\hat{\beta}_{IQ} = 0.0058$$

with

$$\text{Est. S.E.}\left[\hat{\beta}_{IQ} \mid \boldsymbol{X}\right] = \sqrt{\hat{\sigma}^2 S^{IQ}} = 0.001.$$

Hypothesis testing in finite samples: single res. example

Someone claims that each additional IQ point raises one's wage by 0.0075 on average. That is,

$$\beta^0_{IQ} = 0.0075.$$

We want to test this claim. The null and the alternative are

$$\begin{aligned} H_0 &: \beta_{IQ} = 0.0075 \\ H_1 &: \beta_{IQ} \neq 0.0075. \end{aligned}$$

A two-tailed test!

Hypothesis testing in finite samples: single res. example

We need to calculate the t and the t^c, and compare t to t^c to decide on the result of the test.

Hypothesis testing in finite samples: single res. example

t is calculated as follows.

$$t = \frac{0.0058 - 0.0075}{0.001} = -1.75.$$

Hypothesis testing in finite samples: single res. example

t^c is calculated as follows. Consider a significance level of 0.05. Then, for this two-tailed test, $\frac{\alpha}{2} = 0.025$. The degrees of freedom is $935 - 4 = 931$. Then,

$$t^c_{0.025,931} = -1.9625,$$

using the tabulated t distribution at the back of your textbook, or using statistical software.

Hypothesis testing in finite samples: single res. example

Since

$$t > t^c,$$

that is, since

$$-1.7500 > -1.9625,$$

we fail to reject the null hypothesis.

Hypothesis testing in finite samples: single res. example

We can also compare p to p^c to decide on result of the t test. p is the p value corresponding to the t value. p^c is the critical p value corresponding to the critical t value. p^c is called the significance level.

Hypothesis testing in finite samples: single res. example

p is calculated, for this two-tailed test, as

$$p = 2 * p_{-1.75,931} = 0.0805$$

using standard statistical software. The tabulated t distribution at the back of your textbook will not present this exact number because tabulations cannot be too detailed for space reasons.

Hypothesis testing in finite samples: single res. example

p^c is calculated as

$$p^c = 2 * p_{-1.9625,931} = 0.05$$

using standard statistical software, and the tabulated t distribution at the back of your book will present this number because this 0.05 is a conventional critical level.

Hypothesis testing in finite samples: single res. example

Since

$$p > p^c,$$

that is, since

$$0.0805 > 0.0500,$$

we fail to reject the null hypothesis.

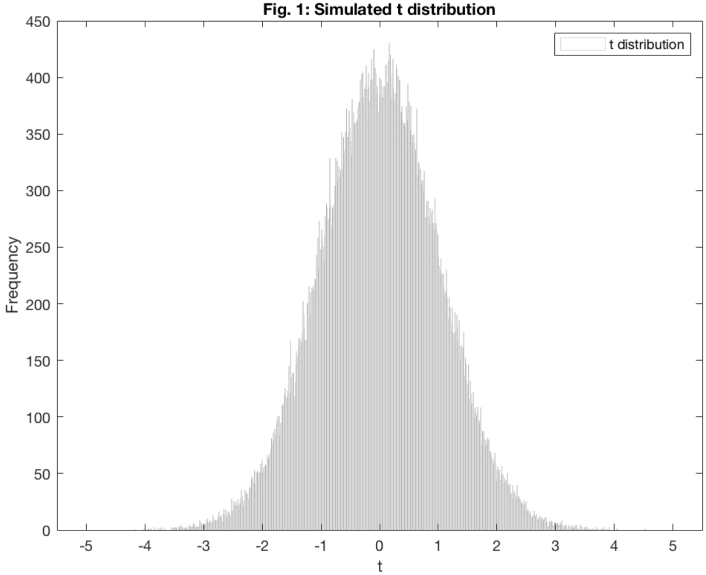
Hypothesis testing in finite samples: single res. example

But calculating the t, t^c, p, p^c as described is complicated, and it is not so clear what we are actually doing.

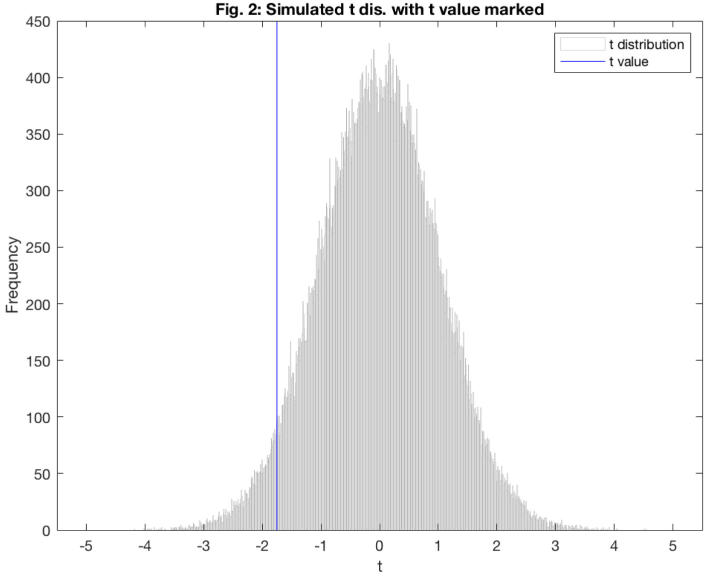
Hypothesis testing in finite samples: single res. example

O'Hara (2018) proposes that econometrics instructors move away from using the tabulated distribution of the test statistic at the back of the textbooks when teaching hypothesis testing. Instead, he proposes that instructors teach students to test hypotheses by using the simulated distribution of the test statistic which can be created using random number generators in statistical software. This provides students with a visual and intuitive understanding of the sampling distribution and the logic behind hypothesis testing. In the next few slides we will follow what O'Hara proposes. You will practice this in the lab.

Hypothesis testing in finite samples: single res. example



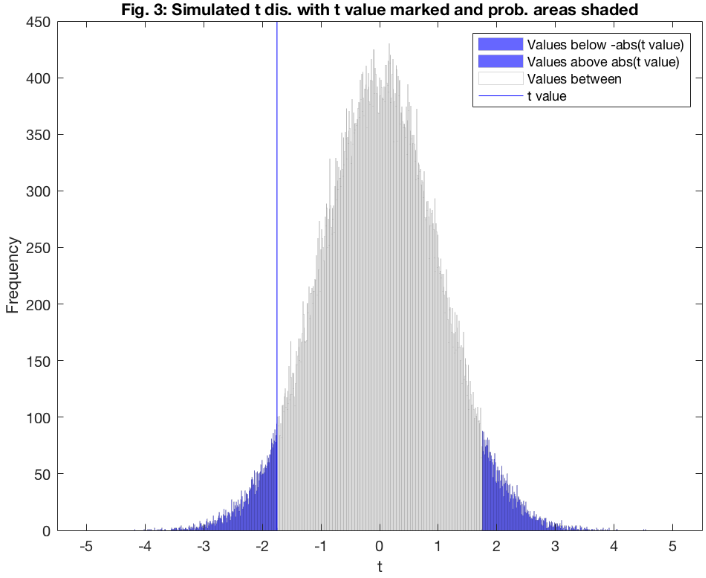
Hypothesis testing in finite samples: single res. example



Hypothesis testing in finite samples: single res. example

Does -1.75 seem like a likely value to observe, if in fact what we have plotted is the true distribution?

Hypothesis testing in finite samples: single res. example



Hypothesis testing in finite samples: single res. example

Consider the shaded area where the values are more extreme than t. The fraction of the extreme t values, among all of the t values, gives the probability area associated with the t value under the simulated t distribution. This gives the simulated p value!

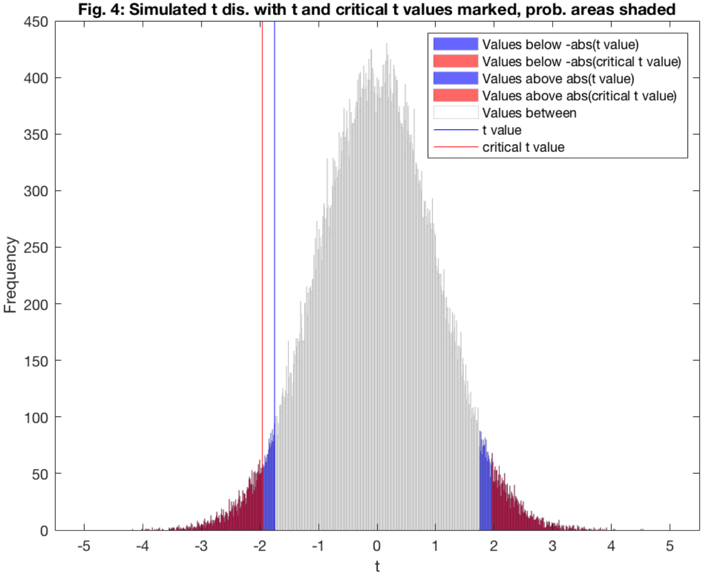
$$p_{sim} = 0.0810.$$

Compare this to

$$p = 0.0805$$

which you would have obtained using the tabulated t distribution. The difference is simulation noise. If you increase the number of draws from the t distribution, p_{sim} will start to converge to p. This is because increasing the number of draws gets the frequency distribution closer to the continuous distribution.

Hypothesis testing in finite samples: single res. example



Hypothesis testing in finite samples: single res. example

Consider the shaded area where the values are more extreme than t^c. The fraction of the extreme values gives the probability area associated with the t^c under the simulated t distribution. This gives the simulated critical p value!

$$p^c_{sim} = 0.0503.$$

Compare this to

$$p^c = 0.0500$$

which you would have obtained using the tabulated t distribution. The difference is simulation noise.

Hypothesis testing in finite samples: single res. example

The whole point of using the simulated distribution of the test statistic while carrying out the hypothesis test is that we have the distribution of the test statistic right in front of us, and we understand what we are actually doing.

Hypothesis testing in large samples

If

$$\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2 \boldsymbol{I}\right]$$

holds, the exact sampling distribution of $\hat{\boldsymbol{\beta}}$, conditional on $\boldsymbol{X}$, is

$$\hat{\boldsymbol{\beta}} \mid \boldsymbol{X} \sim N\left[\boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right].$$

When $\hat{\boldsymbol{\beta}}$ is normal, the t and F statistics have the exact t and F distributions. This is what have shown above.

Hypothesis testing in large samples

If

$$\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2 \boldsymbol{I}\right]$$

does not hold, the t statistic does not have the exact t distribution in finite n. The same holds for the F statistic. Then what happens? Consider the t statistic.

Hypothesis testing in large samples: single restriction

$$
\begin{aligned}
t_k &= \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{\sigma}^2 \left[\left(\boldsymbol{X}'\boldsymbol{X} \right)^{-1} \right]_{k,k}}} \frac{\sqrt{n}}{\sqrt{n}} \\
&= \frac{\sqrt{n} \left(\hat{\beta}_k - \beta_k^0 \right)}{\sqrt{\hat{\sigma}^2 \left[\left(\frac{1}{n} \boldsymbol{X}'\boldsymbol{X} \right)^{-1} \right]_{k,k}}}.
\end{aligned}
$$

Hypothesis testing in large samples: single restriction

Consider the numerator of

$$t_k = \frac{\sqrt{n}\left(\hat{\beta}_k - \beta_k^0\right)}{\sqrt{\hat{\sigma}^2\left[\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}}}.$$

Recall from the derivation of the asymptotic normality of $\hat{\boldsymbol{\beta}}$ that

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left[\mathbf{0}, \sigma^2\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}\right].$$

In this derivation we did not assume that ε is normal! The normal distribution is due to the CLT! Considering the element k of $\hat{\boldsymbol{\beta}}$, and that under the null $\beta_k = \beta_k^0$,

$$\sqrt{n}\left(\hat{\beta}_k - \beta_k^0\right) \xrightarrow{d} N\left[0, \sigma^2\left[\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}\right]_{k,k}\right].$$

Hypothesis testing in large samples: single restriction

Consider the denominator of

$$t_k = \frac{\sqrt{n}\left(\hat{\beta}_k - \beta_k^0\right)}{\sqrt{\hat{\sigma}^2\left[\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}}}.$$

As you have studied in a theoretical exercise

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2.$$

We already know that

$$\left[\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k} \xrightarrow{p} \left[\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}\right]_{k,k}.$$

Using the product rule of plim,

$$\sqrt{\hat{\sigma}^2\left[\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}} \xrightarrow{p} \sqrt{\sigma^2\left[\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}\right]_{k,k}}.$$

Hypothesis testing in large samples: single restriction

$$\sqrt{\hat{\sigma}^2 \left[\left(\frac{1}{n} \boldsymbol{X}' \boldsymbol{X} \right)^{-1} \right]_{k,k}} \xrightarrow{p} \sqrt{\sigma^2 \left[\left(\mathrm{E} \left[\boldsymbol{x}_i \boldsymbol{x}_i' \right] \right)^{-1} \right]_{k,k}}$$

implies

$$\sqrt{\hat{\sigma}^2 \left[\left(\frac{1}{n} \boldsymbol{X}' \boldsymbol{X} \right)^{-1} \right]_{k,k}} \xrightarrow{d} \sqrt{\sigma^2 \left[\left(\mathrm{E} \left[\boldsymbol{x}_i \boldsymbol{x}_i' \right] \right)^{-1} \right]_{k,k}}.$$

Hypothesis testing in large samples: single restriction

Given that

$$\sqrt{n}\left(\hat{\beta}_k - \beta_k^0\right) \xrightarrow{d} N\left[0, \sigma^2\left[\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}\right]_{k,k}\right],$$

and

$$\sqrt{\hat{\sigma}^2\left[\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}} \xrightarrow{d} \sqrt{\sigma^2\left[\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}\right]_{k,k}},$$

and using the ratio rule of limiting distributions (Greene, Theorem D.16),

$$t_k = \frac{\sqrt{n}\left(\hat{\beta}_k - \beta_k^0\right)}{\sqrt{\hat{\sigma}^2\left[\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}}} \xrightarrow{d} N\left[0,1\right].$$

Drop the two instances of $\sqrt{n}$. We have

$$t_k \overset{a}{\sim} N\left[0,1\right].$$

Hypothesis testing in large samples: single restriction

$$t_k \overset{a}{\sim} N[0,1].$$

This shows that the t statistic approximately has a standard normal distribution in finite but large samples. Hence, if n is large, we can compare the t statistic with the critical values from a standard normal distribution. We do not need to assume that ε is normal!

Power of a test

A test is said to have good power if the probability of rejecting the null hypothesis, when it is false, is high.

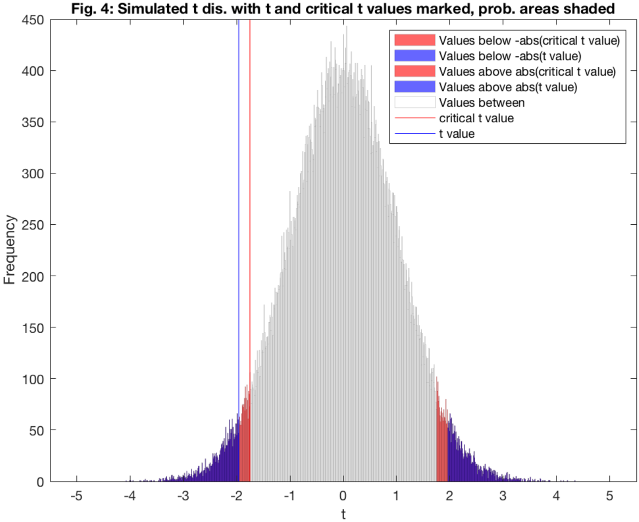
Consider the null and alternative hypotheses

$$H_0 : \beta_k = \beta_k^0$$
$$H_1 : \beta_k \neq \beta_k^0.$$

When is the null hypothesis false? If the alternative is true. When is the alternative true? If you are far on the left or right hand side of the t distribution.

Power of a test

E.g., suppose $\beta_k^0 = 0$. If $-|t_k|$ is smaller than $-|t^c|$, or if $|t_k|$ larger than $|t^c|$, we reject the null.



Power of a test

When are you far on the left or right hand side of the, say, t distribution? In three situations.

Power of a test

First, if the alternative is true. This happens when the effect size is large. Effect size refers to the size of β_k. If β_k is large, then the alternative hypothesis is more likely. β_k is unobserved, but if β_k is large, then the sample we observe is likely to reflect this, and we will have a large $\hat{\beta}_k$, and consequently a large t value.

Power of a test

Second, it happens when the sample size is large. If the sample size is large, the variance of the error is smaller, and consequently the t has a larger value.

Power of a test

Third, if α is larger because this ensures that you are automatically further on the left or on the right of the t distribution. Since everybody agrees on an α value of 5%, this factor does not seem very relevant.

Interval estimation

Is the true coefficient β_k equal to a certain β_k^0? To answer this question we have developed a test statistic, which utilises the data at hand.

Interval estimation

Is the true coefficient β_k within a range of values? To answer this question, we now estimate an interval that β_k could fall in, utilising the data at hand.

Interval estimation

```
. regress wage educ

      Source |       SS           df       MS      Number of obs   =       997
-------------+----------------------------------   F(1, 995)       =    251.46
       Model |  7842.35455         1  7842.35455   Prob > F        =    0.0000
    Residual |  31031.0745       995  31.1870095   R-squared       =    0.2017
-------------+----------------------------------   Adj R-squared   =    0.2009
       Total |   38873.429       996  39.0295472   Root MSE        =    5.5845

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.135645   .0716154    15.86   0.000     .9951106     1.27618
       _cons |  -4.860424   .9679821    -5.02   0.000    -6.759944   -2.960903
------------------------------------------------------------------------------
```

Interval estimation

We know that, in a large sample,

$$t_k = \frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{\sigma}^2 S^{kk}}} \overset{a}{\sim} N\left[0, 1\right].$$

Then,

$$\text{Prob}\left[-1.96 < \frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{\sigma}^2 S^{kk}}} < 1.96\right] = 0.95.$$

if n is large. Here, the choice of the boundaries, and hence the associated probability, is arbitrary.

Interval estimation

Interpret

$$\text{Prob}\left[-1.96 < \frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{\sigma}^2 S^{kk}}} < 1.96\right] = 0.95.$$

The probability that the random variable

$$\frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{\sigma}^2 S^{kk}}}$$

is between the stated boundaries is 0.95.

Interval estimation

Rearrange the terms of

$$\text{Prob}\left[-1.96 < \frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{\sigma}^2 S^{kk}}} < 1.96\right] = 0.95$$

to obtain

$$\text{Prob}\left[\hat{\beta}_k - 1.96\sqrt{\hat{\sigma}^2 S^{kk}} < \beta_k < \hat{\beta}_k + 1.96\sqrt{\hat{\sigma}^2 S^{kk}}\right] = 0.95.$$

Interval estimation

The interpretation is different for

$$\text{Prob}\left[\hat{\beta}_k - 1.96\sqrt{\hat{\sigma}^2 S^{kk}} < \beta_k < \hat{\beta}_k + 1.96\sqrt{\hat{\sigma}^2 S^{kk}}\right] = 0.95.$$

The interpretation is not straightforward. Therefore, we first take account of a couple of points.

Interval estimation

$$\text{Prob}\left[\hat{\beta}_k - 1.96\sqrt{\hat{\sigma}^2 S^{kk}} < \beta_k < \hat{\beta}_k + 1.96\sqrt{\hat{\sigma}^2 S^{kk}}\right] = 0.95.$$

First, the interpretation is for the unique nonrandom population parameter β_k.

Interval estimation

$$\text{Prob}\left[\hat{\beta}_k - 1.96\sqrt{\hat{\sigma}^2 S^{kk}} < \beta_k < \hat{\beta}_k + 1.96\sqrt{\hat{\sigma}^2 S^{kk}}\right] = 0.95.$$

Second, the end points of the interval are random because $\hat{\beta}_k$ is random. $\hat{\beta}_k$ has a sampling distribution. We are taking samples from the population repeatedly, and calculating an interval using each sample. Hence, we have a series of intervals resulting from repeated sampling. But since we are not able to do repeated sampling, we estimate the interval using the data at hand.

Interval estimation

$$\text{Prob}\left[\hat{\beta}_k - 1.96\sqrt{\hat{\sigma}^2 S^{kk}} < \beta_k < \hat{\beta}_k + 1.96\sqrt{\hat{\sigma}^2 S^{kk}}\right] = 0.95.$$

The interpretation is as follows. In repeated sampling, the true population parameter β_k falls 95 percent of the times within the intervals like the one we estimated using the data at hand.

Interval estimation

Given the single sample at hand, we have only one estimate of the interval. The probability that the interval we estimate using the data at hand contains β_k is either 0 or 1. Hence, it is incorrect to say that the probability that the interval we estimated using the data at hand contains β_k is 95 percent. The interval we calculated is just an estimate of one of the many intervals that contain β_k 95 percent of the times.

Interval estimation, example

A test and a confidence interval are closely related.

```
. regress wage educ

      Source |       SS           df       MS      Number of obs   =       997
-------------+----------------------------------   F(1, 995)       =    251.46
       Model |  7842.35455         1  7842.35455   Prob > F        =    0.0000
    Residual |  31031.0745       995  31.1870095   R-squared       =    0.2017
-------------+----------------------------------   Adj R-squared   =    0.2009
       Total |   38873.429       996  39.0295472   Root MSE        =    5.5845

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.135645   .0716154    15.86   0.000     .9951106     1.27618
       _cons |  -4.860424   .9679821    -5.02   0.000    -6.759944   -2.960903
------------------------------------------------------------------------------
```

We reject the null $\beta_{educ} = 0$ of the t test since it lies outside the confidence interval.

Hypothesis testing in finite samples: multiple restrictions

We might want to hypothesise that there are J linear restrictions on the true coefficient vector $\boldsymbol{\beta}$ against alternatives such that

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

$\boldsymbol{R}$ is a matrix of J restrictions for K parameters. $J \times K$. $\boldsymbol{\beta}$ is the true coefficient vector. $K \times 1$. $\boldsymbol{q}$ is the hypothesised value of $\boldsymbol{R}\boldsymbol{\beta}$. $J \times 1$. E.g.,

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{R}} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}}_{\boldsymbol{\beta}} = \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{\boldsymbol{q}},$$

implies the linear restrictions

$$\begin{aligned} \beta_1 + \beta_2 &= 0, \\ \beta_2 &= 1. \end{aligned}$$

Hypothesis testing in finite samples: multiple restrictions

We want to test whether our hypothesis is true. We do not observe $\boldsymbol{\beta}$, but we can estimate it with $\hat{\boldsymbol{\beta}}$. Suppose

$$\boldsymbol{R}\hat{\boldsymbol{\beta}} = \boldsymbol{q}.$$

We could conclude that H_0 is true. But this conclusion has a problem.

Hypothesis testing in finite samples: multiple restrictions

Remember that $\hat{\boldsymbol{\beta}}$ is a random variable and has a sampling distribution. Hence, there is a probability associated with the condition

$$\boldsymbol{R}\hat{\boldsymbol{\beta}} = \boldsymbol{q}.$$

Therefore, we need to check if

$$\boldsymbol{R}\hat{\boldsymbol{\beta}} = \boldsymbol{q}$$

holds in statistical terms. We do this check through a test statistic based on the random variable

$$\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}.$$

Hypothesis testing in finite samples: multiple restrictions

$\boldsymbol{R}\hat{\beta} - \boldsymbol{q}$ is a random variable, and therefore it has a distribution. Therefore, we start by studying this distribution.

Hypothesis testing in finite samples: multiple restrictions

Taking the expectation conditional on $\boldsymbol{X}$,

$$\begin{aligned}
\mathrm{E}\left[\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q} \mid \boldsymbol{X}\right] &= \mathrm{E}\left[\boldsymbol{R}\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right] - \mathrm{E}\left[\boldsymbol{q} \mid \boldsymbol{X}\right] \\
&= \boldsymbol{R}\mathrm{E}\left[\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right] - \boldsymbol{q} \\
&= \boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{q} \\
&= \mathbf{0},
\end{aligned}$$

assuming that

$$\mathrm{E}\left[\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right] = \boldsymbol{\beta},$$

and under the null

$$\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{q}.$$

Hypothesis testing in finite samples: multiple restrictions

Taking the variance conditional on $\boldsymbol{X}$,

$$\begin{aligned}
\text{Var}\left[\boldsymbol{R}\hat{\beta} - \boldsymbol{q} \mid \boldsymbol{X}\right] &= \text{Var}\left[\boldsymbol{R}\hat{\beta} \mid \boldsymbol{X}\right] \\
&= \boldsymbol{R}\text{Var}\left[\hat{\beta} \mid \boldsymbol{X}\right]\boldsymbol{R}' \\
&= \boldsymbol{R}\sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}' \\
&= \sigma^2\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'.
\end{aligned}$$

Hypothesis testing in finite samples: multiple restrictions

We know that if

$$\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2 \boldsymbol{I}\right]$$

then

$$\hat{\boldsymbol{\beta}} \mid \boldsymbol{X} \sim N\left[\boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right].$$

Since $\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}$ is a linear function of $\hat{\boldsymbol{\beta}}$,

$$\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2 \boldsymbol{R} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{R}'\right],$$

with the mean and variance derived above.

Hypothesis testing in finite samples: multiple restrictions

$$\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2 \boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right].$$

Then,

$$\left[\sigma^2 \boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right]^{-1/2}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right) \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \boldsymbol{I}\right],$$

by Greene, Theorem B.10. Then,

$$W \mid \boldsymbol{X} = \left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right)'\left[\sigma^2 \boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right) \mid \boldsymbol{X} \sim \chi^2\left[J\right],$$

by Greene, Theorem B.11. The distribution does not depend on the model parameters $\hat{\boldsymbol{\beta}}$, σ, or on the data $\boldsymbol{X}$. Hence,

$$W \sim \chi^2\left[J\right].$$

Hypothesis testing in finite samples: multiple restrictions

$$W = \frac{\left(\boldsymbol{R}\hat{\beta} - \boldsymbol{q}\right)' \left[\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\hat{\beta} - \boldsymbol{q}\right)}{\sigma^2}.$$

W is not usable due to the unknown σ^2. Replace σ^2 with its unbiased estimator

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-K}.$$

Hypothesis testing in finite samples: multiple restrictions

Consider further manipulations to lead W to become F which will follow a known distribution. That is, consider

$$
\begin{aligned}
F &= \frac{\left(\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{q}\right)'\left[\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{q}\right)}{\hat{\sigma}^2}\frac{1}{J}\frac{\sigma^2}{\sigma^2}\frac{n-K}{n-K} \\
&= \frac{\left(\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{q}\right)'\left[\sigma^2\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{q}\right)/J}{\left(\hat{\sigma}^2/\sigma^2\right)(n-K)/(n-K)}.
\end{aligned}
$$

Hypothesis testing in finite samples: multiple restrictions

$$F = \frac{\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right)' \left[\sigma^2 \boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right)/J}{\left(\hat{\sigma}^2/\sigma^2\right)(n-K)/(n-K)}.$$

As shown above, the numerator is distributed as

$$\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right)' \left[\sigma^2 \boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right) = W \sim \chi^2[J].$$

It can be shown that the term in the denominator is distributed as

$$\left(\hat{\sigma}^2/\sigma^2\right)(n-K) \sim \chi^2[n-K].$$

It can be shown that these two terms are orthogonal (Greene, p. 158). Then,

$$F \sim \frac{\chi^2[J]/J}{\chi^2[n-K]/(n-K)} = F[J, n-K].$$

Hypothesis testing in finite samples: multiple restrictions

$$F = \frac{\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right)' \left[\sigma^2 \boldsymbol{R} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{R}'\right]^{-1} \left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right) / J}{\left(\hat{\sigma}^2/\sigma^2\right)(n-K)/(n-K)}.$$

Having shown that the F statistic has a F distribution, drop the appearances of σ^2 and $n-K$ to obtain

$$F = \frac{\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right)' \left[\boldsymbol{R}\hat{\sigma}^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{R}'\right]^{-1} \left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right)}{J}.$$

Hypothesis testing in finite samples: multiple restrictions

$$F = \frac{\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right)' \left[\boldsymbol{R}\hat{\sigma}^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right]^{-1} \left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right)}{J}.$$

What is the intuition of the test statistic? Is the distance between $\boldsymbol{R}\hat{\boldsymbol{\beta}}$ and $\boldsymbol{q}$ sufficiently large, with the distance measured in terms of the sampling variance of $\boldsymbol{R}\hat{\boldsymbol{\beta}}$? Is F sufficiently large? If it is, reject the null that $\boldsymbol{R}\hat{\boldsymbol{\beta}} = \boldsymbol{q}$. This is the decision rule of the test.

Hypothesis testing in finite samples: multiple restrictions

How large should F be depends on the threshold F value we want to consider. This threshold, or critical, F value is

$$F^c_{1-\alpha,J,n-K}.$$

F^c: a value from the F distribution which depends on the following.

$1-\alpha$: area under the F distribution covering up to where we want F^c to rest. Hence, $1-\alpha$ determines the threshold F^c we want to consider.

J, $n-K$: two degrees of freedom parameters as arguments of the F distribution.

Hypothesis testing in finite samples: multiple restrictions

This means that we would reject the null hypothesis if

$$F > F^{c}_{1-\alpha,J,n-K}.$$

If the null is rejected, we conclude that the restrictions we impose on the parameters in the null hypothesis are jointly not significant. The test does not inform about which restriction is not significant: any or all restrictions are not significant.

Hypothesis testing in finite samples: multiple res. example

Mincer (1974) considers the regression of the log of wage on exper, educ (in years), and IQ score. The data contains 935 observations. Add to this regression age and age squared.

.

Hypothesis testing in finite samples: multiple res. example

Someone claims that age has no effect on wage. That is,

$$\beta^0_{age} = 0,\ \beta^0_{age\ squared} = 0.$$

We want to test this claim. The null and the alternative hypotheses are

$$H_0 : \boldsymbol{R}\hat{\boldsymbol{\beta}} = \boldsymbol{q}$$
$$H_1 : \boldsymbol{R}\hat{\boldsymbol{\beta}} \neq \boldsymbol{q}.$$

Checkpoint. Note that this is a one-tailed test.

Hypothesis testing in finite samples: multiple res. example

We need to calculate F, and compare it to F^c to decide on the result of the test.

Hypothesis testing in finite samples: multiple res. example

F should be calculated as described above. Checkpoint. It turns out that

$$F = 4.5735.$$

Hypothesis testing in finite samples: multiple res. example

Consider a significance level of 0.05. Hence, for this one-tailed test, $\alpha = 0.05$. The numerator degrees of freedom is 2, and the denominator degrees of freedom is $935 - 6 = 929$. F^c can then be calculated as

$$F^c = 3.0054,$$

using the tabulated F distribution at the back of your book, or using statistical software.

Hypothesis testing in finite samples: multiple res. example

Since

$$F > F^c,$$

that is, since

$$4.5735 > 3.0054,$$

we reject the null hypothesis.

Hypothesis testing in finite samples: multiple res. example

We can also compare p to p^c to decide on result of the F test. p is the p value corresponding to the F value. p^c is the critical p value corresponding to the critical F value.

Hypothesis testing in finite samples: multiple res. example

p can be calculated as

$$p = p_{4.5735,2,929} = 0.0106$$

using standard statistical software, and the tabulated F distribution at the back of your book will not present this number because tabulations cannot be too detailed for space reasons.

Hypothesis testing in finite samples: multiple res. example

p^c can be calculated as

$$p^c = p_{3.0054,2,929} = 0.0500$$

using standard statistical software, and the tabulated t distribution at the back of your book will present this number because 0.05 is a conventional critical level.

Since

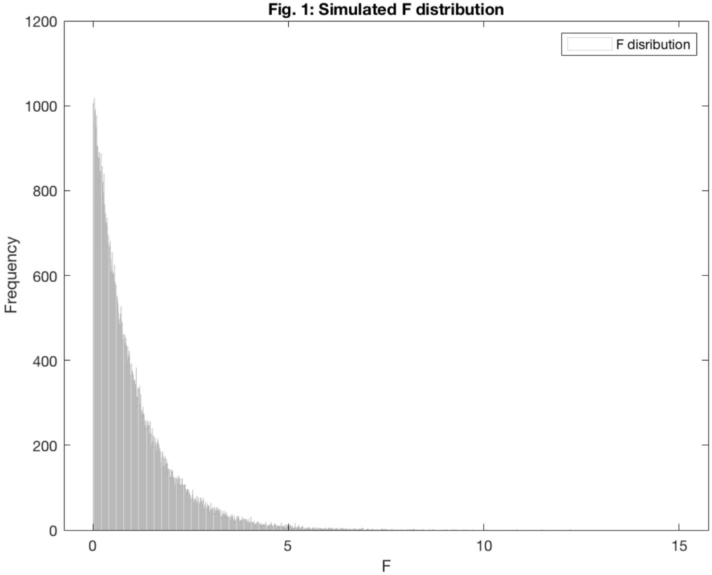
$$p > p^c,$$

that is, since

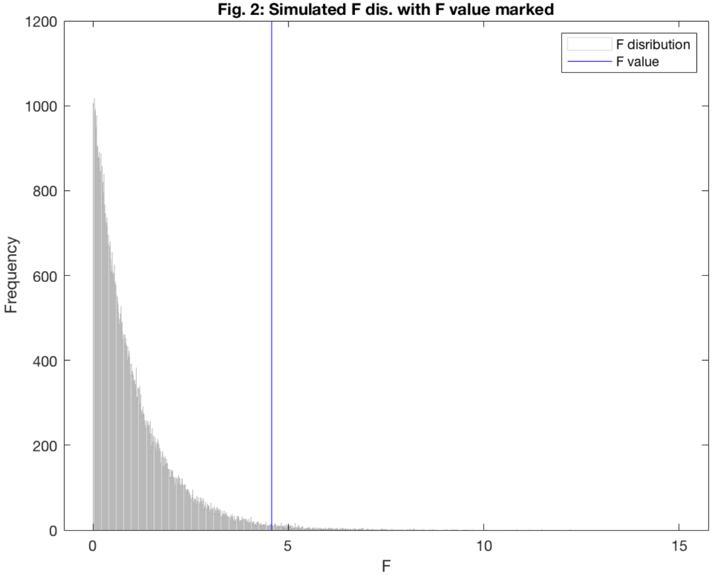
$$0.0106 < 0.0500,$$

we reject the null hypothesis.

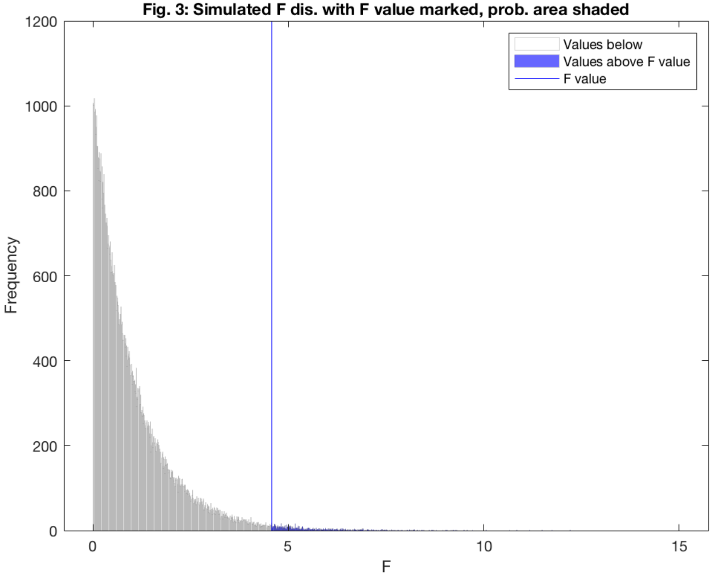
Hypothesis testing in finite samples: multiple res. example



Hypothesis testing in finite samples: multiple res. example



Hypothesis testing in finite samples: multiple res. example



Hypothesis testing in finite samples: multiple res. example

Consider the shaded area where the values are more extreme than F. The fraction of the extreme values gives the probability area associated with the F value under the simulated F distribution. This gives the simulated p value!

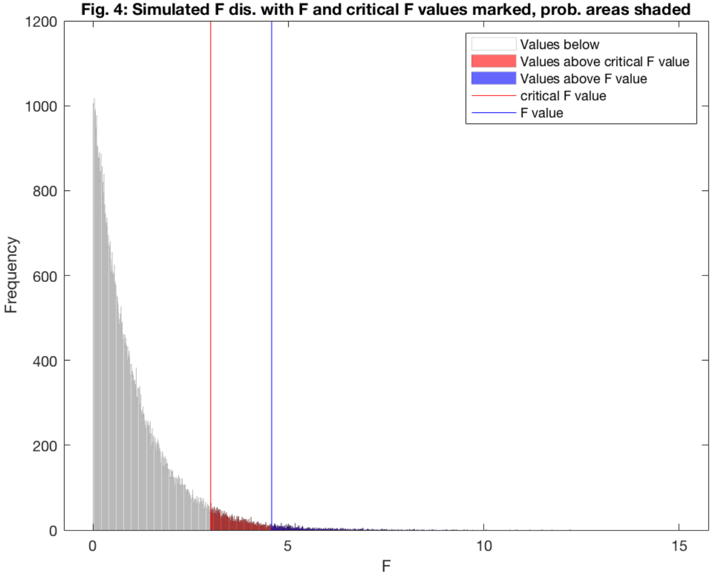
$$p_{sim} = 0.0101.$$

Compare this to

$$p = 0.0106$$

which you would have obtained using the tabulated F distribution. The difference is due to simulation noise.

Hypothesis testing in finite samples: multiple res. example



Hypothesis testing in finite samples: multiple res. example

Consider the shaded area where the values are more extreme than F^c. The fraction of the extreme values gives the probability area associated with the F^c value under the simulated F distribution. This gives the simulated critical p value!

$$p^c_{sim} = 0.0504.$$

Compare this to

$$p^c = 0.0500$$

which you would have obtained using the tabulated t distribution. The difference is due to simulation noise.