Interval estimation

Econometrics for minor Finance, Lecture 5

Tunga Kantarcı, Fall 2025

Interval estimation

We tested whether a true coefficient

$$\beta$$

is equal to a certain value

$$\beta^0$$

in a probabilistic term. That is, utilizing the sample data at hand, we have point estimated β using the OLS method, and developed a test statistic to check how close $\hat{\beta}$ and β^0 are in a statistical sense.

Interval estimation

We can also estimate a lower and upper bound for the true coefficient

$$\beta$$

That is, utilizing the sample data at hand, we can construct an interval estimate.

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 finite sample,

$$t = \frac{\hat{\beta} - \beta^0}{\mathsf{SEE}\left[\hat{\beta}\right]} \sim t\left[n - K\right]$$

Interval estimation

Then, we can state that

$$\text{Prob}\left(-t_{\alpha/2,\nu} < \frac{\hat{\beta} - \beta^0}{\text{SEE}\left[\hat{\beta}\right]} < t_{\alpha/2,\nu}\right) = 1 - \alpha$$

where

- α is some probability value, and
- $-t_{\alpha/2}$ and $t_{\alpha/2}$ are some lower and upper thresholds, or critical values as we have seen.

Interval estimation

Interpret

$$\mathsf{Prob}\left(-t_{\alpha/2,\nu} < \frac{\hat{\beta} - \beta^0}{\mathsf{SEE}\left[\hat{\beta}\right]} < t_{\alpha/2,\nu}\right) = 1 - \alpha$$

The probability that the random variable

$$\frac{\hat{\beta} - \beta^0}{\mathsf{SEE}\left[\hat{\beta}\right]}$$

is between the stated thresholds is $1 - \alpha$.

Interval estimation

For example, if $\nu = 999$ and $\alpha = 0.05$,

$$t_{0.025,999} = 1.9623,$$

and hence

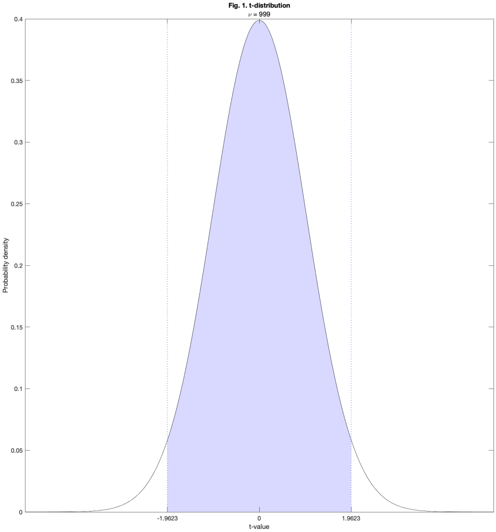
$$\text{Prob}\left(-1.9623 < \frac{\hat{\beta} - \beta^0}{\text{SEE}\left[\hat{\beta}\right]} < 1.9623\right) = 0.95$$

The probability that the random variable

$$\frac{\hat{\beta} - \beta^0}{\text{SEE}\left[\hat{\beta}\right]}$$

is between the stated boundaries is 0.95. This the shaded area between the stated thresholds is 0.95 in the figure below.

Interval estimation



Interval estimation

Now rearrange the terms of

$$\text{Prob}\left(-t_{\alpha/2,\nu} < \frac{\hat{\beta} - \beta^0}{\text{SEE}\left[\hat{\beta}\right]} < t_{\alpha/2,\nu}\right) = 1 - \alpha$$

to obtain

$$\text{Prob}\left(\hat{\beta} - t_{\alpha/2,\nu}\text{SEE}\left[\hat{\beta}\right] < \beta^0 < \hat{\beta} + t_{\alpha/2,\nu}\text{SEE}\left[\hat{\beta}\right]\right) = 1 - \alpha$$

Interval estimation: Intermezzo: Notation switch

We have

$$\text{Prob}\left(\hat{\beta} - t_{\alpha/2,\nu}\text{SEE}\left[\hat{\beta}\right] < \beta^0 < \hat{\beta} + t_{\alpha/2,\nu}\text{SEE}\left[\hat{\beta}\right]\right) = 1 - \alpha$$

The interval is about the true parameter β, so we rewrite

$$\text{Prob}\left(\hat{\beta} - t_{\alpha/2,\nu}\text{SEE}\left[\hat{\beta}\right] < \beta < \hat{\beta} + t_{\alpha/2,\nu}\text{SEE}\left[\hat{\beta}\right]\right) = 1 - \alpha$$

Interval estimation: Intermezzo: Notation switch

In the test, we fix a null value β^0 and check whether $\hat{\beta}$ is compatible with it. In interval estimation, we change the perspective. Rather than testing a single hypothesized value, we ask which values of β would be compatible with the observed $\hat{\beta}$. The algebra is the same, but the role of the symbol changes, from a single null value

$$\beta^0$$

to the set of plausible parameter values

$$\beta$$

Interval estimation

At this instance the interpretation changes.

$$\text{Prob}\left(\hat{\beta} - t_{\alpha/2,\nu}\text{SEE}\left[\hat{\beta}\right] < \beta < \hat{\beta} + t_{\alpha/2,\nu}\text{SEE}\left[\hat{\beta}\right]\right) = 1 - \alpha$$

Notice two things.

Interval estimation

$$\text{Prob}\left(\hat{\beta} - t_{\alpha/2,\nu}\text{SEE}\left[\hat{\beta}\right] < \beta < \hat{\beta} + t_{\alpha/2,\nu}\text{SEE}\left[\hat{\beta}\right]\right) = 1 - \alpha$$

First, the interpretation is for the unique nonrandom population parameter β^0.

Interval estimation

$$\text{Prob}\left(\hat{\beta} - t_{\alpha/2,\nu}\text{SEE}\left[\hat{\beta}\right] < \beta < \hat{\beta} + t_{\alpha/2,\nu}\text{SEE}\left[\hat{\beta}\right]\right) = 1 - \alpha.$$

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

Interval estimation

$$\text{Prob}\left(\hat{\beta} - t_{\alpha/2,\nu}\text{SEE}\left[\hat{\beta}\right] < \beta < \hat{\beta} + t_{\alpha/2,\nu}\text{SEE}\left[\hat{\beta}\right]\right) = 1 - \alpha$$

Then, the interpretation is as follows. In repeated sampling, the true population parameter β^0 falls within intervals, like the one we estimate using the data at hand, $1 - \alpha$ of the times.

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

$$\beta$$

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

$$\beta$$

is 95 percent. The interval we calculated is just an estimate of one of the many intervals that contain

$$\beta$$

95 percent of the times.

Interval estimation: Example

Building on the earlier example, if $\nu = 999$ and $\alpha = 0.05$,

$$t_{0.025,999} = 1.9623$$

Suppose

$$\hat{\beta} = 0.4574$$

and

$$\mathrm{SEE}\left[\hat{\beta}\right] = 0.0557$$

We get

$$\mathrm{Prob}\left(0.3482 < \beta < 0.5667\right) = 95\%$$

So an interval estimate using the sample data at hand is

$$[0.3482, 0.5667]$$

This is called a "confidence" interval because we use this, and only one interval, to be confident about the population coefficient to a certain probability extent.

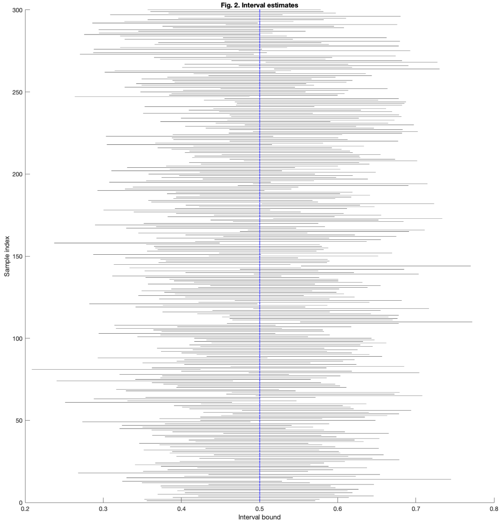
Interval estimation

The figure below shows that

$$\beta$$

falls within all intervals 95 percent of the times.

Interval estimation



Interval estimation: Example

```
. 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: Example

A test and a confidence interval are closely related. We reject the null

$$\beta_{educ} = 0$$

of the t test since it lies outside the confidence interval.

Interval estimation

$$\text{Prob}\left(\hat{\beta} - z_{\alpha/2}\,\text{SEE}\left[\hat{\beta}\right] < \beta < \hat{\beta} + z_{\alpha/2}\,\text{SEE}\left[\hat{\beta}\right]\right) = 1 - \alpha$$

In large samples, the t distribution is well approximated by the standard normal distribution. Thus, we use

$$z_{\alpha/2}$$

instead of

$$t_{\alpha/2,\nu}$$

The interpretation remains the same.