Confidence Intervals and Regions

1 Exact confidence intervals

Hypothesis tests are the foundation for all inference in classical econometrics. We can find how compatible economics data and theory is with one another through these hypothesis tests. We often wish to make inference on our parameter estimates. To do this, we could perform a battery of tests. Alternatively, we construct a confidence interval. Confidence intervals may be either exact or approximates. When the exact distribution of a test statistic is known, the coverage is equal to the confidence interval, and the interval is exact. Otherwise, we may have to be content with approximate confidence intervals.

Definition: The probability that the random interval includes or covers the time value of the parameter is called the coverage probability or just the coverage.

Like a confidence interval, a $1-\alpha$ confidence region for a set of k-model parameters, such as the components of a k-vector parameter θ, is a region in k-dimensional space, such that for every point in the vector θ_0 in the confidence region, the joint hypothesis that $\theta = \theta_0$ is not rejected by the appropriate member of a family of tests at level α.

Suppose $H_0 : \theta = \theta_0$, and $\mathbf{y}$ was our sample space and $r(\mathbf{y}, \theta_0)$ was our test statistic for the null hypothesis. Then

$$\Pr_{\theta_0}(r(y, \theta_0) \leq c_\alpha) = 1 - \alpha,$$

where c_α is defined as the critical value of the distribution of $r(y, \theta_0)$ under the null, and the confidence interval

$$\{\theta_0 \in \Theta : r(\mathbf{y}, \theta_0) \leq c_\alpha\}$$

corresponds to all values of θ_0 such that $H_0 : \theta = \theta_0$ is not rejected.

Definition: If the finite sample distribution of the test is known, we obtain an exact confidence interval.

Definition: If, as is more often the case, only the test statistic's asymptotic distribution is known, we attain an asymptotic confidence interval.

We shall soon see what or how we are able to construct a confidence interval by "inverting" a test statistic.

Example: Consider the hypothesis $H_0 : \beta_2 = \beta_2^0$, where

$$r(y, \beta_{20}) = \frac{\hat{\beta}_2 - \beta_2^0}{s_{\beta_2}} \sim t(n-k), \tag{1}$$

under the null hypothesis with a given probability:

$$\Pr\left(t_{\frac{\alpha}{2}} \leq \frac{\beta_2 - \beta_2^0}{s_{\beta_2}} \leq t_{1-\frac{\alpha}{2}}\right) = 1 - \alpha,$$

the symmetric confidence can be derived as follows

$$\left\{\beta_2^0 \in \mathbb{R} : t_{\frac{\alpha}{2}} \leq \frac{\beta_2 - \beta_2^0}{s_{\beta_2}} \leq t_{1-\frac{\alpha}{2}}\right\}$$

$$\begin{aligned}
\implies t_{\frac{\alpha}{2}} &\leq \frac{\hat{\beta}_2 - \beta_2^0}{s_{\beta_2}} \leq t_{1-\frac{\alpha}{2}} \\
= \underbrace{\hat{\beta}_2 - t_{1-\frac{\alpha}{2}} s_{\beta_2}}_{\text{not dependent on } \beta_2^0} &\leq \beta_2^0 \leq \underbrace{\hat{\beta}_2 - t_{\frac{\alpha}{2}} s_{\beta_2}}_{\text{not dependent on } \beta_2^0},
\end{aligned}$$

and so the $1 - \alpha\%$ [symmetric] confidence interval is

$$\left(\hat{\beta}_2 - t_{1-\frac{\alpha}{2}} s_{\beta_2}, \hat{\beta}_2 - t_{\frac{\alpha}{2}} s_{\beta_2}\right).$$

1.1 Exact confidence regions

When we are interested in making inferences about the values of two or more parameters, it can be quite misleading to look at the confidence intervals for each of the parameters individually. By using confidence intervals, we are implicitly basing our inferences on the marginal distributions of the parameter estimates.

The confidence intervals we have obtained thus far are from inverting t tests based on families of statistics of the form $\left(\hat{\theta} - \theta_0\right)/s_{\theta_0}$. If we wish instead to construct a confidence region, we must invert joint tests for several parameters. These are usually tests based on statistics that follow the F or χ^2 distributions.

Example: We construct the confidence region for $\boldsymbol{\beta}_{k\times 1}$ by first considering the hypothesis

$$H_0 : \ \boldsymbol{\beta} = \boldsymbol{\beta}_0,$$

and the following test statistic

$$r(\mathbf{y}, \boldsymbol{\beta}_0) = \frac{\left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top (\mathbf{X}^\top \mathbf{X})^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right] / k}{\hat{\sigma}^2} \sim F(k, n-k).$$

The $1-\alpha\%$ confidence region is then given by

$$\left\{ \forall \boldsymbol{\beta}_0 : \frac{\left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top (\mathbf{X}^\top \mathbf{X})^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right] / k}{\hat{\sigma}^2} \leq c_\alpha \right\},$$

where c_α is the $1-\alpha$ quantile of the $F(k, n-k)$ distribution. Note here, also, that c_α does not depend on $\boldsymbol{\beta}_0$.

2 Asymptotic confidence intervals

Confidence intervals may be either exact or approximate. The confidence intervals and regions we just derived have been exact. When the exact distribution of the test statistics used to construct a confidence interval is known, the coverage will be equal to the confidence level, and the interval will be exact. Otherwise, we may have to be content with approximate confidence intervals, which may be based either on asymptotic theory or on bootstrapping.

Consider the following

$$\sqrt{n}(\hat{\beta}_2 - \beta_2^0) \overset{a}{\sim} N(0, \sigma^2 \mathbf{S}^{-1}_{\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2}), \tag{2}$$

where

$$\begin{aligned}
\mathbf{S}_{\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2} &= \mathbf{S}_{\mathbf{X}_2^\top \mathbf{X}_2} - \mathbf{S}_{\mathbf{X}_2^\top \mathbf{X}_1} \mathbf{S}^{-1}_{\mathbf{X}_1^\top \mathbf{X}_1} \mathbf{S}_{\mathbf{X}_1^\top \mathbf{X}_2}, \\
n^{-1} \mathbf{X}_2^\top \mathbf{X}_2 &\xrightarrow{p} \mathbf{S}_{\mathbf{X}_2^\top \mathbf{X}_2}, \\
n^{-1} \mathbf{X}_2^\top \mathbf{X}_1 &\xrightarrow{p} \mathbf{S}_{\mathbf{X}_2^\top \mathbf{X}_1}, \\
n^{-1} \mathbf{X}_1^\top \mathbf{X}_1 &\xrightarrow{p} \mathbf{S}_{\mathbf{X}_1^\top \mathbf{X}_1}, \\
n^{-1} \mathbf{X}_1^\top \mathbf{X}_2 &\xrightarrow{p} \mathbf{S}_{\mathbf{X}_1^\top \mathbf{X}_2}, \\
n^{-1} \mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2 &\xrightarrow{p} \mathbf{S}_{\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2}.
\end{aligned}$$

Like our asymptotic hypothesis testing, under fairly weak conditions, we see that the $\hat{\beta}_2$ is asymptotically normal, or exhibits asymptotic normality. What (2) is saying is that the asymptotic variance of $\sqrt{n}(\hat{\beta}_2 - \beta_2^0)$ is the limit of

$\sigma^2(n^{-1}\mathbf{X}_2^\top\mathbf{M}_1\mathbf{X}_2)^{-1}$ as $n \to \infty$. In practice, we divide this by n and use $s^2(\mathbf{X}_2^\top\mathbf{M}_1\mathbf{X}_2)^{-1}$ to estimate $\text{Var}(\hat{\beta}_2)$, where s^2 is the usual OLS estimate of the error variance. Thus, we can yield an asymptotic [approximate] confidence interval, by first writing

$$\frac{\sqrt{n}(\hat{\beta}_2 - \beta_2^0)}{\sqrt{s^2\left(\frac{\mathbf{X}_2^\top\mathbf{M}_1\mathbf{X}_2}{n}\right)^{-1}}} \overset{a}{\sim} N(0,1),$$

which gives

$$\left(\hat{\beta}_2 - z_{1-\frac{\alpha}{2}} s_{\hat{\beta}_2}, \hat{\beta}_2 - z_{\frac{\alpha}{2}} s_{\hat{\beta}_2}\right).$$

2.1 Asymptotic confidence regions

Asymptotic confidence regions are a natural extension of the above confidence intervals. Consider the following:

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \overset{a}{\sim} N(0, \sigma^2\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}^{-1}), \tag{3}$$

where, as usual,

$$n^{-1}\mathbf{X}^\top\mathbf{X} \xrightarrow{p} \mathbf{S}_{\mathbf{X}^\top\mathbf{X}}.$$

Whenever $n^{-1}\mathbf{X}^\top\mathbf{X}$ tends to $\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}$ as $n \to \infty$, the matrix $(\mathbf{X}^\top\mathbf{X})^{-1}$, without the factor of n, simply tends to a zero matrix. This is just a consequence of the fact that $\hat{\boldsymbol{\beta}}$ is consistent. (3) gives us the rate of convergence of $\hat{\boldsymbol{\beta}}$ to its probability limit of $\boldsymbol{\beta}_0$. Since multiplying the estimation error by $\sqrt{n}$ gives rise to an expression of zero mean and finite covariance matrix, it follows that the estimation error itself tends to zero at the same rate as $n^{-\frac{1}{2}}$. This property is expressed by saying that the estimator $\hat{\boldsymbol{\beta}}$ is root-n consistent.

Quite generally, let $\hat{\boldsymbol{\theta}}$ be a root-n consistent, asymptotically normal, estimator of a parameter vector $\boldsymbol{\theta}$. Any estimator of the covariance matrix of $\hat{\boldsymbol{\theta}}$ must tend to zero as $n \to \infty$. Let $\boldsymbol{\theta}_0$ denote the true value of $\boldsymbol{\theta}$, and let $\mathbf{V}$ denote the limiting covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$. Then, an estimator $\hat{\text{Var}}(\hat{\boldsymbol{\theta}})$ is said to be a consistent estimator of the covariance matrix of $\hat{\boldsymbol{\theta}}$ if

$$\operatorname*{plim}_{n\to\infty} (n\hat{\text{Var}}(\hat{\boldsymbol{\theta}})) = \mathbf{V}.$$

Thus, if $\hat{\boldsymbol{\beta}}$ is root-n consistent and asymptotically normal, and if $\hat{\text{Var}}(\hat{\boldsymbol{\beta}})$ is a consistent estimator of the variance of $\hat{\boldsymbol{\beta}}$, then we can write

$$\begin{aligned}
(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top (\hat{\text{Var}}(\hat{\boldsymbol{\beta}}))^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &= \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top (n\hat{\text{Var}}(\hat{\boldsymbol{\beta}}))^{-1} \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \qquad (4) \\
&= \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \left[s^2 \left(\frac{\mathbf{X}^\top \mathbf{X}}{n} \right)^{-1} \right]^{-1} \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&= (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \frac{\mathbf{X}^\top \mathbf{X}}{s^2} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)
\end{aligned}$$

$$\implies (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \frac{\mathbf{X}^\top \mathbf{X}}{s^2} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \overset{a}{\sim} \chi^2(k). \qquad (5)$$

Since $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically normal under the null, with mean zero, and since the middle factor of the above expression tends to the inverse of its limiting covariance matrix, (4) is of the form $\mathbf{x}^\top \boldsymbol{\Omega}^{-1} \mathbf{x}$, and so (5) is asymptotically distributed under a null hypothesis as $\chi^2(k)$. Thus, we have the following confidence region

$$\text{CR} \sim \left\{ \forall \boldsymbol{\beta}_0 : (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \frac{\mathbf{X}^\top \mathbf{X}}{s^2} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \leq c_\alpha \right\},$$

where c_α is the $1 - \alpha$ quantile of the $\chi^2(k)$ distribution.

2.2 Bootstrap Confidence Interval

When exact confidence intervals are not available, asymptotic ones are normally used. However, just as asymptotic tests do not always perform well in finite samples, neither do asymptotic confidence intervals. We know bootstrapped P-values generally perform better than their asymptotic counterparts, so it seems natural, then, to base confidence intervals on bootstrap tests when asymptotic tests give poor coverage.

When we wish to construct a bootstrap confidence interval, we must first use a bootstrap DGP that satisfies the null hypothesis. It may appear that we must use an infinite number of bootstrap DGP's if we are to consider the full family of tests, each with a different null. Thankfully, we can avoid that.

It is vital for a bootstrap DGP to satisfy the null hypothesis that is being tested, in order to construct a bootstrap test. However, when the distribution of the test statistic does not depend on precisely which null is being tested, the same bootstrap distribution can be used for a whole family of tests with different nulls.

Suppose that we wish to construct a bootstrap confidence interval based on the t-statistic:

$$\hat{t}(\theta_0) = r(y, \theta_0) = \frac{\hat{\theta} - \theta_0}{SE(\hat{\theta})}$$

The first step is to compute $\hat{\theta}$ and its standard error using the original data, $\mathbf{y}$. Then we generate bootstrap samples using a DGP, which may be either parametric or semiparametric, characterised by $\hat{\theta}$ and by other relevant estimates, such as the error variance (which may be needed). The resulting bootstrap DGP is thus quite independent of θ_0 but it does depend on the estimate $\hat{\theta}$.

We can generate B bootstrap samples, y_j^*, $j = 1, ..., B$. For each of these, we compute an estimate θ_j* and its standard error $SE(\theta_j)^*$ like we do with the original data, and then we compute the bootstrap "t statistic":

$$t_i^* = r(y_j^*, \hat{\theta}) = \frac{\theta_j - \hat{\theta}}{SE(\theta_j)^*}$$

Repeat this step B times. Then sort the B bootstrapped t statistics from lowest to highest. Once that's done, select $c_{\frac{\alpha}{2}}^*$ and $c_{1-\frac{\alpha}{2}}^*$ from the sorted bootstrapped t-statistics and form the following confidence interval:

$$\left(\hat{\theta} - c_{1-\frac{\alpha}{2}}^* SE(\hat{\theta}), \hat{\theta} + c_{\frac{\alpha}{2}}^* SE(\hat{\theta})\right). \tag{6}$$