Toulouse School of Economics

M1 INTERMEDIATE ECONOMETRICS

Hypothesis testing

Koen Jochmans François Poinas

2024 — 2025

Overview

This deck of slides goes through hypothesis testing in the context of the linear regression model.

The corresponding chapter in Hansen is 9.

Framework (H9.2-H9.3)

Let

$$\theta = r(\beta)$$

for $r : \mathbb{R}^k \to \mathbb{R}^q$.

Wish to test a null hypothesis

$$\mathbb{H}_0 : \theta = \theta_0$$

against the alternative

$$\mathbb{H}_1 : \{\theta : \theta \neq \theta_0\}.$$

Do this by looking at the sample, via a statistic T.

Use a decision rule:

Accept the null if $T \leq c$,

Reject the null if $T > c$,

for a chosen critical value c.

Examples (H9.5)

The simplest case has θ a single linear combination of the regression vector β:

$$\theta = r'\beta$$

for some vector r.

r could be an element of the standard basis for $\mathbb{R}^k$; this picks out particular elements of $\beta = (\beta_1, \ldots, \beta_k)'$.

In the classical linear regression model, for this case, T would be the t-statistic.

For testing multiple linear combinations, $\theta = \boldsymbol{R}'\beta$ in the classical linear regression model one usually uses the F-statistic.

These approaches do not extend beyond the classical linear regression setup.

Type I and II errors (H9.4 and H9.6)

The statistic T is random.

So the decision rule is subject to statistical error.

Type I error: Reject the null when it is true.

Type II error: Accept the null when it is false.

The size is the probability of a type I error.

If the distribution of T is known under the null the critical value can be chosen as to control size exactly.

The t-test and F-test in the classical linear regression model are again examples.

In general, the distribution of T (under the null) is unknown, so we rely on asymptotic approximations instead.

Wald test (H9.10-H9.11)

From before we know that

$$n\,(\hat{\theta}-\theta)'\hat{\boldsymbol{V}}_{\theta}^{-1}(\hat{\theta}-\theta) \underset{d}{\to} \chi^2_q.$$

Here, θ is unknown.

Under the null $\theta = \theta_0$ is known, and so

$$W = n\,(\hat{\theta}-\theta_0)'\hat{\boldsymbol{V}}_{\theta}^{-1}(\hat{\theta}-\theta_0) \underset{d}{\to} \chi^2_q.$$

Let G_q be the CDF of the χ^2_q-distribution. Then

$$\mathbb{P}(W > c \,|\, \mathbb{H}_0 \text{ is true}) \to 1 - G_q(c)$$

as $n \to \infty$.
A test with (asymptotic) size $\alpha \in (0,1)$ is obtained on setting

$$c = G_q^{-1}(1-\alpha),$$

the $1-\alpha$ quantile of G_q.

p-value (H9.8)

Note that

$$c = G_q^{-1}(1-\alpha)$$

is decreasing in α. So for a decision rule with a smaller size, the critical value threshold is higher.

The (asymptotic) p-value is

$$p = 1 - G_q(W).$$

This is the (asymptotic) probability of observing a test statistic at least as large as the one observed in the data.

Also gives the minimal α at which one would reject the null hypothesis.

t-statistic

When θ is a scalar

$$W = T^2$$

for

$$T = \sqrt{n}\,\hat{V}_\theta^{-1/2}(\hat{\theta} - \theta_0) \underset{d}{\rightarrow} N(0,1),$$

which is the t-statistic.

The decision rules to reject the null when

$$W > G_q^{-1}(1-\alpha)$$

or

$$|T| > \Phi^{-1}(1 - \alpha/2)$$

are equivalent.

We call this a t-test by convention, but its distribution is not Student t!

F-statistic

Suppose that we have homoskedastic errors. Then we would use the non-robust variance estimator

$$\hat{\boldsymbol{V}}_{\beta}^{0} = s^2\, \hat{\boldsymbol{Q}}_{XX}^{-1}$$

with $s^2 = (\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}})/(n-k)$.

In the classical linear regression model, for $\theta = \boldsymbol{R}'\beta$, a null of the form $\theta = \theta_0$ is tested using the F-statistic

$$F = W^0/q$$

which has an exact $F_{q,n-k}$-distribution there.

No longer true in our setting. The statistic never follows and F-distribution.

Now, $q\,F \underset{d}{\rightarrow} \chi^2_q$.

In the classical setting,

$$Y|X = x \sim N(x'\beta, \sigma^2).$$

In this case,

$$\begin{pmatrix} \hat{\beta} - \beta \\ \hat{\boldsymbol{e}} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e} \\ \boldsymbol{M}\boldsymbol{e} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \\ \boldsymbol{M} \end{pmatrix} \boldsymbol{e}$$

and so

$$\left.\begin{pmatrix} \hat{\beta} - \beta \\ \hat{\boldsymbol{e}} \end{pmatrix}\right| \boldsymbol{X} \sim N\left(0, \begin{pmatrix} \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} & 0 \\ 0 & \sigma^2\boldsymbol{M} \end{pmatrix}\right)$$

because, for the off-diagonal block, we have

$$\mathbb{E}(\boldsymbol{e}'\boldsymbol{M}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e}) = \sigma^2 \text{trace}(\boldsymbol{X}'\boldsymbol{M}(\boldsymbol{X}'\boldsymbol{X})^{-1}) = 0,$$

which, by normality, implies independence.

For any linear combination $\theta = \boldsymbol{R}'\beta$, estimated by $\hat{\theta} = \boldsymbol{R}'\hat{\beta}$, we then equally have

$$\left(\begin{array}{c} \hat{\theta} - \theta \\ \hat{\boldsymbol{e}} \end{array} \right) \Bigg| \boldsymbol{X} \sim N\left(0, \left(\begin{array}{cc} \sigma^2 \boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R} & 0 \\ 0 & \sigma^2 \boldsymbol{M} \end{array} \right)\right)$$

Furthermore, with $\boldsymbol{M} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'$ the eigendecomposition and $\boldsymbol{u} = \boldsymbol{H}'\boldsymbol{e}$,

$$\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e} = \boldsymbol{e}'\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\boldsymbol{e} = \boldsymbol{u}'\boldsymbol{\Lambda}\boldsymbol{u}.$$

Because $\boldsymbol{M}$ is a projection matrix of rank $n-k$, the diagonal of $\boldsymbol{\Lambda}$ contains k zeros and $(n-k)$ ones. Let $\boldsymbol{u}_1$ be the subset of $\boldsymbol{u}$ associated with the $(n-k)$ non-zero eigenvalues. Then

$$\boldsymbol{u}_{\mathbf{1}} \sim N(0, \sigma^2 \boldsymbol{I}_{n-k}), \qquad \boldsymbol{e}'\boldsymbol{M}\boldsymbol{e} = \boldsymbol{u}_{\mathbf{1}}'\boldsymbol{u}_{\mathbf{1}} \sim \sigma^2 \chi^2_{n-k}.$$

Therefore,

$$(n-k)\frac{s^2}{\sigma^2} = \frac{\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}}{\sigma^2} \sim \chi^2_{n-k}.$$

Let θ be univariate. Then

$$\frac{\hat{\theta} - \theta}{\sigma \sqrt{r'(\boldsymbol{X}'\boldsymbol{X})^{-1} r}} \sim N(0, 1),$$

and so

$$\left(\frac{\hat{\theta} - \theta}{\sigma \sqrt{r'(\boldsymbol{X}'\boldsymbol{X})^{-1} r}} \right) \Bigg/ \sqrt{\frac{s^2}{\sigma^2}} = \frac{\hat{\theta} - \theta}{s \sqrt{r'(\boldsymbol{X}'\boldsymbol{X})^{-1} r}} \sim t_{n-k}.$$

Let θ be q-variate. Then

$$\frac{(\hat{\theta} - \theta)'(\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}(\hat{\theta} - \theta)}{\sigma^2} \sim \chi^2_q$$

and so

$$\left(\frac{(\hat{\theta} - \theta)'(\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}(\hat{\theta} - \theta)}{\sigma^2} \right) \Bigg/ q \, \frac{s^2}{\sigma^2} = W^0/q \sim F_{q,n-k}.$$

Consistency (H9.22)

Let

$$\beta(\theta) = \mathbb{P}(\text{reject } \mathbb{H}_0 | \theta)$$

be the power function.

For our Wald statistic,

$$\beta(\theta) = \mathbb{P}(W > c \,|\, \theta).$$

Note that $\lim_{n\to\infty} \beta(\theta_0) = \alpha$.

A test is consistent if

$$\lim_{n\to\infty} \beta(\theta) = 1$$

for any fixed value $\theta \neq \theta_0$.

For a t-statistic

$$\begin{aligned} T &= \sqrt{n}\,\hat{V}_\theta^{-1/2}(\hat{\theta} - \theta_0) \\ &= \sqrt{n}\,\hat{V}_\theta^{-1/2}(\hat{\theta} - \theta) + \sqrt{n}\,\hat{V}_\theta^{-1/2}(\theta - \theta_0) \\ &\underset{d}{\to} N(0,1) + \sqrt{n}\,V_\theta^{-1/2}(\theta - \theta_0). \end{aligned}$$

Under the null, $\theta = \theta_0$ and so $T \underset{d}{\to} N(0,1)$ as $n \to \infty$.

For any $\theta \neq \theta_0$, the second term grows to $+\infty$ or $-\infty$ at the rate $\sqrt{n}$ as $n \to \infty$.

Consequently, for $h_\theta = V_\theta^{-1/2}(\theta - \theta_0)$ and $Z \sim N(0,1)$, the power function of the t-test with size α satisfies

$$\begin{aligned} \beta(\theta) \to\ & \mathbb{P}(|Z + \sqrt{n}h_\theta| > z_{1-\alpha/2} \,|\, \theta) \\ &= \mathbb{P}(Z + \sqrt{n}h_\theta \ < -z_{1-\alpha/2} \,|\, \theta) + \mathbb{P}(Z + \sqrt{n}h_\theta \ > \ z_{1-\alpha/2} \,|\, \theta) \\ &= \Phi(-z_{1-\alpha/2} - \sqrt{n}h_\theta) + 1 - \Phi(z_{1-\alpha/2} - \sqrt{n}h_\theta) \to 1 \end{aligned}$$

as $n \to \infty$.

For a Wald statistic the same reasoning goes through as

$$W \underset{d}{\rightarrow} \chi_q^2 + 2\sqrt{n}\, Z' h_\theta + n\, h_\theta' h_\theta,$$

where, now,

$$Z \sim N(0, I_q), \qquad h_\theta = \boldsymbol{V}_\theta^{-1/2}(\theta - \theta_0).$$

The last term, $n h_\theta' h_\theta$, is non-negative and grows to $+\infty$ as $n \to \infty$.

Consistency does not tell us whether a test is powerful in practice.

Does not allow to compare different test statistics.

As $n \to \infty$, any fixed alternative lies increasingly far from the null.

We can approximate the power by keeping the alternative close to the null as n grows.

Do this by considering a sequence of alternatives:

$$\theta_n = \theta_0 + n^{-1/2} h$$

for some fixed h.

For the t-test we now get

$$\begin{aligned} T &= \sqrt{n}\,\hat{V}_\theta^{-1/2}(\hat{\theta} - \theta_0) \\ &= \sqrt{n}\,\hat{V}_\theta^{-1/2}(\hat{\theta} - \theta_n) + \sqrt{n}\,\hat{V}_\theta^{-1/2}(\theta_n - \theta_0) \underset{d}{\to} N(0,1) + V_\theta^{-1/2} h \end{aligned}$$

and so

$$T \underset{d}{\to} N(\delta, 1), \qquad \delta = V_\theta^{-1/2} h,$$

which is a normal distribution that is not centered at zero.

Then

$$\lim_{n\to\infty} \beta(\theta_n) = \Phi(-z_{1-\alpha/2} - \delta) + 1 - \Phi(z_{1-\alpha/2} - \delta) = \pi(\delta)$$

is the asymptotic local power function. (This implicitly depends on the chosen size α.)

As $h = \sqrt{n}(\theta_n - \theta_0)$ and $V_\theta^{1/2}$ is the asymptotic standard deviation of its estimator $\sqrt{n}(\hat{\theta} - \theta_0)$, δ is a relative measure that reflects how large the violation of the null is compared to the noise in the estimator. As $|\delta| \to \infty$ then $\pi(\delta) \to 1$.

To approximate actual power for a given n and alternative θ, we solve $\theta = \theta_0 + h/\sqrt{n}$ for h to find $h = \sqrt{n}(\theta - \theta_0)$. Our power approximation then is

$$\pi\left(\sqrt{n}\, V^{-1/2}(\theta - \theta_0)\right).$$

For the Wald test we similarly, get that

$$W \underset{d}{\rightarrow} \chi^2_q(h' \boldsymbol{V}_\theta^{-1} h)$$

which is a non-central χ^2_q-distribution, wit non-centrality parameter $h' \boldsymbol{V}_\theta^{-1} h$.

Confidence intervals (H7.13 and H7.18 and H9.20)

Consider again the scalar case first.

Because

$$\sqrt{n}\,(\hat{\theta} - \theta) \underset{d}{\rightarrow} N(0, V_\theta)$$

as $n \to \infty$, for any $\alpha \in (0,1)$,

$$\mathbb{P}\left(\sqrt{n}\hat{V}_\theta^{-1/2}(\hat{\theta} - \theta) \leq \Phi^{-1}(\alpha)\right) \to \alpha$$

or, equivalently,

$$\mathbb{P}\left(\hat{\theta} \leq \theta + \sqrt{\frac{\hat{V}_\theta}{n}}\,\Phi^{-1}(\alpha)\right) \to \alpha.$$

Take any $\alpha < {}^1\!/_2$, then

$$\mathbb{P}\left(\hat{\theta} \leq \theta + \sqrt{\frac{\hat{V}_\theta}{n}}\,\Phi^{-1}(1 - {}^{\alpha}\!/_{2})\right) - \mathbb{P}\left(\hat{\theta} \leq \theta + \sqrt{\frac{\hat{V}_\theta}{n}}\,\Phi^{-1}({}^{\alpha}\!/_{2})\right)$$

equals

$$\mathbb{P}\left(\theta + \sqrt{\frac{\hat{V}_\theta}{n}}\,\Phi^{-1}({}^{\alpha}\!/_{2}) < \hat{\theta} \leq \theta + \sqrt{\frac{\hat{V}_\theta}{n}}\,\Phi^{-1}(1 - {}^{\alpha}\!/_{2})\right) \to 1 - \alpha.$$

Re-arranging and exploiting that $\Phi^{-1}(1 - {}^{\alpha}\!/_{2}) = -\Phi^{-1}({}^{\alpha}\!/_{2})$ gives the equivalence

$$\mathbb{P}\left(\hat{\theta} + \sqrt{\frac{\hat{V}_\theta}{n}}\,\Phi^{-1}(1 - {}^{\alpha}\!/_{2}) > \theta \geq \hat{\theta} - \sqrt{\frac{\hat{V}_\theta}{n}}\,\Phi^{-1}(1 - {}^{\alpha}\!/_{2})\right) \to 1 - \alpha.$$

The interval

$$C = \left[\hat{\theta} - \sqrt{\frac{\hat{V}_\theta}{n}}\,\Phi^{-1}(1 - {}^{\alpha}\!/\!_{2}), \sqrt{\frac{\hat{V}_\theta}{n}}\,\Phi^{-1}(1 - {}^{\alpha}\!/\!_{2})\right]$$

covers the parameter θ with probability $1 - \alpha$ (in large samples).

Letting

$$T(\theta) = \frac{\hat{\theta} - \theta}{\sqrt{\hat{V}_\theta / n}},$$

we equivalently have that

$$\mathbb{P}\left(-\Phi^{-1}(1 - {}^{\alpha}\!/\!_{2}) < T(\theta) \leq \Phi^{-1}(1 - {}^{\alpha}\!/\!_{2})\right) \to 1 - \alpha$$

That is, we can write

$$C = \{\theta^* : |T(\theta^*)| \leq \Phi^{-1}(1 - {}^{\alpha}\!/\!_{2})\}.$$

This is known as 'inversion' of a test statistic to construct a confidence interval.

In the vector case, we can construct a confidence region based on the Wald statistic.

In this case, if we take

$$C = \{\theta^* : W(\theta^*) \leq G_q^{-1}(1-\alpha)\},$$

we obtain

$$\mathbb{P}(\theta \in C) \to 1 - \alpha$$

as $n \to \infty$.

Remember that the probability works on the set C, not on θ (which is a fixed constant).