General Framework of hypothesis tests

- define a vector (matrix) R that picks up one element of a vector (matrix):

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \qquad R = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

we will see that $R\beta = \beta_2$

$$R\,\beta = \beta_2$$

$$\begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \beta_2$$

hence: if we want to test $H_0: \beta_2 = 0$, we can write

$$H_0: \underbrace{R\,\beta}_{\beta_2 \text{ is selected}} = 0$$

now, let's define a scalar (vector r)

$$R\beta = r$$

where $r = 0$, i.e

$$\begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = 0$$

$$R \cdot \beta = r$$

now, we can rewrite $\underline{H_0}$: $R\beta = r$ against $R\beta \neq r$.

for a single coeff, we use a t.stat.

$$\hat{var}(b_2) = \hat{var}(Rb) = R\left(\hat{var}(b)\right)R'$$

$$\begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \underline{\hat{var}\ b_1} & \hat{cov}\ b_1 b_2 & \hat{cov}\ b_1 b_3 \\ \hat{cov}\ b_1 b_2 & \underline{\hat{var}\ b_2} & \hat{cov}\ b_2 b_3 \\ \hat{cov}\ b_3 b_1 & \hat{cov}\ b_3 b_2 & \underline{\hat{var}\ b_3} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

This multiplication PICKS UP $\hat{var}(b_2)$

since $t = \dfrac{b_2 - \beta_2^\circ}{\sqrt{\hat{var}\, b_2}}$, we can write:

$$t = \frac{Rb - R\beta}{\sqrt{\underbrace{R\, \hat{var}(b)\, R'}_{\text{scalar}}}} = \frac{Rb - r}{\sqrt{R\, \hat{\sigma}^2 (X'X)^{-1} R'}}$$

$$= \frac{Rb - r}{\hat{\sigma} \sqrt{R(X'X)^{-1}R'}}$$

and $\quad Rb \sim N\left[R\beta,\ \sigma^2 R(X'X)^{-1}R'\right]$

FOR a NUMBER J OF LINEAR HYPOTHESES WE HAVE:

$H_0: R\beta = r \qquad H_A: R\beta \neq r$

$(J \times K)(K \times 1) = (J \times 1)$

and we use an F statistic, because it is a joint test of J hypotheses:

$$F = \frac{(Rb - r)' \left[R \, \widehat{\text{var}}(b) \, R' \right]^{-1} (Rb - r)}{J}$$

$$F = \frac{(Rb - r)' \left[R \, (X'X)^{-1} R' \right]^{-1} (Rb - r)}{J \cdot \hat{\sigma}^2}$$

IS DISTRIBUTED $\boxed{F_{J, (T-K)}}$ under H_0

We reject $H_0 : R\beta = r$ is $F > F_c(\alpha)$ at level α

ex: in the model $y_t = \beta_1 + \beta_2 p_t + \beta_3 a_t + \beta_4 a_t^2 + e_t$

$$H_0: \beta_1 + 2\beta_2 + 40\beta_3 + 1600\beta_4 = 175$$
$$\beta_3 + 80\beta_4 = 1$$

$$R = \begin{bmatrix} 1 & 2 & 40 & 1600 \\ 0 & 0 & 1 & 80 \end{bmatrix}, \quad r = \begin{bmatrix} 175 \\ 1 \end{bmatrix}$$

you can see that easily when you rewrite the hypothesis:

$$\begin{bmatrix} 1\beta_1 + 2\beta_2 + 40\beta_3 + 1600\beta_4 \\ 0\cdot\beta_1 + 0\cdot\beta_2 + 1\cdot\beta_3 + 80\cdot\beta_4 \end{bmatrix} \begin{matrix} = \\ = \end{matrix} \begin{bmatrix} 175 \\ 1 \end{bmatrix}$$

hence

$$Rb - r = \begin{bmatrix} 1 & 2 & 40 & 1600 \\ 0 & 0 & 1 & 80 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} - \begin{bmatrix} 175 \\ 1 \end{bmatrix}$$

THE RESTRICTED LEAST SQUARES

b_*

Introducing non-sample information improves the efficiency of estimators (reduces the variance)

. it can be done by reparametrizing the model

• or by using the Restricted Least squares estimator the information is embodied in

$$H_0: \; R\beta = r$$

If H_0 is not rejected, one may wish to reestimate the model, incorporating the restriction = H_0 $R\beta = r$.

A reason to do it is to improve the efficiency. This restricted estimation produces estimator b_* which satisfies $Rb_* = r$

how to derive b_*:

$$y = X\beta + e$$

define a scalar function of squared errors from the restricted estimation

$$\hat{e}'\hat{e} = \Sigma \hat{e}_t^{*2}$$

$$\varphi = (y - Xb_*)'(y - Xb_*) - 2\lambda'(Rb_* - r)$$

Where λ denotes a column vector of **J** Lagrange multipliers. Taking partial derivatives:

$$\frac{\partial \varphi}{\partial b_*} = -2X'y + 2X'Xb_* - 2R'\lambda = 0 \text{, at min}$$

$$\frac{\partial \varphi}{\partial \lambda} = -2(Rb_* - r) = 0 \text{, at min}$$

We minimize the sum of squares subject to a set of restrictions: constrained minimization. From the above equation:

$$X'Xb_* - X'y - R'\lambda = 0$$

$$Rb_* - r = 0$$

Premultiply the first equation by $R(X'X)^{-1}$ yielding

$$* \quad Rb_* - R(X'X)^{-1}X'y - R(X'X)^{-1}R'\lambda = 0$$

using $Rb_* - r = 0$ and knowing that $b = (X'X)^{-1}X'y$

we can solve for λ:

$$\lambda = \left[R(X'X)^{-1}R'\right]^{-1}(r - Rb)$$

substituting into *

$$b_* = (X'X)^{-1}X'y + (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}(r - Rb)$$

i.e.

$$b_* = b + (X'X)^{-1}R'\left[R(X'X)^{-1}R\right]^{-1}(r - Rb)$$

where b is the unrestricted OLS estimator $(X'X)^{-1}X'y$

this formula defines the restricted least squares estimator satisfying the set of J restrictions in $Rb_* = r$.

Define now the RESIDUAL VECTOR:

$$
\begin{aligned}
\hat{e}_* &= y - Xb_* \\
&= \underbrace{y - Xb} - X(b_* - b) \\
&= \hat{e} - X(b_* - b)
\end{aligned}
$$

where $\hat{e}$ is the OLS residual from non-restricted regression

transpose and multiply:

$$\hat{e}_*'\hat{e}_* = \hat{e}'\hat{e} + (b_* - b)'X'X(b_* - b)$$

the cross product vanishes since $X'\hat{e} = 0$. Hence, the difference btw the restricted and unrestricted residual sums of squares can be written:

$$\hat{e}_*'\hat{e}_* - \hat{e}'\hat{e} = (\underline{b_* - b})'X'X(\underline{b_* - b})$$

from the formula : $b_* = b + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - Rb)$

$\underline{b_* - b} = (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - Rb)$. substitute:

$$\hat{e}_*'\hat{e}_* - \hat{e}'\hat{e} = (r - Rb)'[R(X'X)^{-1}R']^{-1}R\underline{(X'X)^{-1}\cancel{X'X}}(X'X)^{-1}$$
$$R'[R(X'X)^{-1}R']^{-1}(r - Rb)$$

finaly,

$$\boxed{\hat{e}_*'\hat{e}_* - \hat{e}'\hat{e} = (r - Rb)'\left[R(X'X)^{-1}R'\right]^{-1}(r - Rb)}$$

and is always ≥ 0

compare now to the expression for F test of a set of linear restrictions: (unrestricted regression)

$$F = \frac{1}{J}(Rb - r)'\left[R\,\widehat{cov}(b)\,R'\right]^{-1}(Rb - r)$$

we know that $\widehat{cov}(b) \equiv \widehat{var}(b)$ (notation) $= \hat{\sigma}^2(X'X)^{-1}$

$$= \frac{1}{J}\left(\frac{1}{\hat{\sigma}^2}\right)\underbrace{(Rb - r)'\left[R(X'X)^{-1}R'\right]^{-1}(Rb - r)}$$

$$F = \frac{1}{J}\,\frac{1}{\hat{\sigma}^2}\left(\hat{e}_*'\hat{e}_* - \hat{e}'\hat{e}\right)$$

and since $\hat{\sigma}^2 = \hat{e}'\hat{e}/(T-K)$, we have

$$\boxed{F = \frac{(\hat{e}_*'\hat{e}_* - \hat{e}'\hat{e})/J}{\hat{e}'\hat{e}\ /\ T-K}}$$

$$F = \frac{(RSS_R - RSS_U)/J}{RSS_U / T-K}$$

PREDICTION IN A GENERAL LINEAR MODEL

$$y_t = \beta_1 + x_{t2}\beta_2 + x_{t3}\beta_3 + x_{t4}\beta_4 + x_{t5}\beta_5 + e_t$$

$e_t \sim N(0, \sigma^2)$

given: $x_0 = (1 \;\; x_{02} \;\; x_{03} \;\; x_{04} \;\; x_{05})'$ we predict y_0:

$$y_0 = \beta_1 + x_{02}\beta_2 + x_{03}\beta_3 + x_{04}\beta_4 + x_{05}\beta_5 + e_0$$

$$= x_0'\beta + e_0$$

the best linear predictor:

$$\boxed{\hat{y}_0 = x_0' b}$$

where $b = (X'X)^{-1}X'y$, the OLS estimator. 2 sources of error of prediction: b is estimator of β and e_0 is not necessarily $= 0$.

variance of the prediction error:

$$\mathrm{var}(\hat{y}_0 - y_0) = \sigma^2 \left[1 + x_0' (X'X)^{-1} x_0 \right]$$

replace σ^2 by $\hat{\sigma}^2$

$$\frac{\hat{y}_0 - y_0}{\sqrt{\widehat{\mathrm{var}}(\hat{y}_0 - y_0)}} \sim t_{(T-K)}$$

for hypothesis test and prediction interval building:

$$\left[\hat{y}_0 - t_c \cdot \sqrt{\widehat{\mathrm{var}}(\hat{y}_0 - y_0)} \, , \; \hat{y}_0 + t_c \sqrt{\widehat{\mathrm{var}}(\hat{y}_0 - y_0)} \right]$$

is a prediction interval for $\hat{y}_0$.

COLLINEAR VARIABLES

MULTICOLLINEARITY (variables "move" together

one var is a linear function of the other).

Symptom – very large variances of individual coeff, low t statistic, but jointly all regressors signif.

ASYMPTOTIC TESTS

IN THE GENERAL LINEAR MODEL

OLS estimator: $b = \beta + (X'X)^{-1}X'e$

has a distribution that depends on the distribution of error e.
IF X is RANDOM, IT DEPENDS ON THE DISTRIBUTION OF X TOO

CASE 1: X NONRANDOM, e NORMALLY DISTRIBUTED

we have an exact result: $b \sim N(\beta, \sigma^2(X'X)^{-1})$

and hence t-distributed t-ratio, Fisher distributed F test, ect.

CASE 2:

EITHER X NONRANDOM, e NOT NORMAL
OR X RANDOM, e NORMAL
OR X RANDOM, e NOT NORMAL

we use asymptotic test.

- exception: X and e normaly distributed → b is CONDITIONALLY NORMAL, and CONDITIONAL t and F

WE HAVE TO USE ASYMPTOTIC METHODS, i.e:

- ASYMPTOTICALY VALID ESTIMATORS
- ASYMPT. VALID TESTS

ASSUMPTIONS REQUIRED:

1) $\left(\frac{X'X}{T}\right)$ converges to a finite nonsingular matrix Σ_{xx}

2) the random X is at least contemporaneously uncorrelated with e.

Then, by CENTRAL LIMIT THEOREM:

$$\frac{b_K - \beta_K}{\sqrt{\hat{var}(b_K)}} \overset{A}{\sim} N(0,1)$$

i.e. as $T \to \infty$ "asymptotically normally distributed"

note: this suggest using a normal distribution rather that t for hypo test. But if sample is small, you can still use t, even in large samples as well, for testing a hypothesis on a single parameter: $H_0: \beta_K = \beta$

IN TESTING JOINT HYPO WE WILL USE χ^2 DISTRIBUTIC

$$H_0: R\beta = r \quad \text{agains} \quad H_1: R\beta \neq r$$

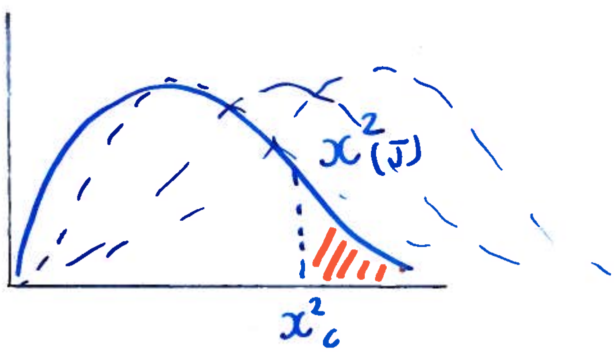
R is $(J \times K)$, r is $(J \times 1)$ for J constrainst on K parameters

THE WALD TEST

asymptotically valid test for testing any linear hypotheses, jointly:

$$\lambda_W = \frac{(Rb-r)'\,[R(X'X)^{-1}R']^{-1}(Rb-r)}{\hat{\sigma}^2}$$

$$= \frac{RSS_R - RSS_U}{\hat{\sigma}^2} \overset{A}{\sim} \chi^2_{(J)}$$



reject $H_0: Rb = r$ at level $\alpha = 5\%$

if $\lambda_W > \chi^2_c$

note: $\hat{\sigma}^2 = RSS_U/(T-K)$ OR $\hat{\sigma}^2 = RSS_U/T$

THE LAGRANGE MULTIPLIER TEST

$$\lambda_M = \frac{(Rb-r)'\,[R(X'X)^{-1}R']^{-1}(Rb-r)}{\hat{\sigma}_*^2}$$

$$= \frac{RSS_R - RSS_U}{\hat{\sigma}_*^2} \overset{A}{\sim} \chi^2_{(J)}$$

note: $\sigma_*^2 = RSS_R/(T-K-J)$ OR $\sigma_*^2 = RSS_R/T$

however, you can still use the F test in finite samples, it is not wrong in practice:

$$F = \frac{\lambda_W}{J} = \frac{(RSS_R - RSS_U)/J}{\hat{\sigma}^2} \sim F_{(J,\, T-K)}$$

THE LIKELIHOOD RATIO TEST

applies to the MLE.

$L(H_0)$ maximized likelihood under H_0

$L(H_1)$ under the alternative

$$\lambda_{LR} = 2\left[L(H_A) - L(H_0)\right] \overset{A}{\sim} \chi^2_{(J)}$$

reject H_0 when $\lambda_{LR} > \chi^2_c$

if the restricted maximum is <u>much</u> less than the unrestricted then we believe that the alternative is true and this is evidence against, i.e. to reject H_0.

in our linear model and under normal errors,

$$\lambda_{LR} = T\,(\ln RSS_R - \ln RSS_U)$$

HYPOTHESES TESTS

1. SMALL SAMPLE

- linear model
- the classical assumptions are supposed to hold

ex: $y_t = \beta_1 + \beta_2 p_t + \beta_3 a_t + e_t$, $e_t \sim N(0, \sigma^2)$

1.1. TESTS OF HYPOTHESES AVAILABLE BY DEFAULT FROM SOFTWARE

a) $H_0: \beta_j = 0$ against $H_A: \beta_j \neq 0$

$$t = \frac{b_j}{s.e.(b_j)} = \frac{b_j}{\sqrt{\hat{var}\, b_j}}$$

for $j = 1, 2, \ldots K$

under H_0 , $t \sim t_{(T-K)}$

b) H_0 : "all slopes $= 0$" against H_A "at least one slope is not zero" : for my ex: $H_0: \beta_2 = \beta_3 = 0$

<u>unrestricted model:</u>

$$y_t = b_1 + b_2 p_t + b_3 a_t + \hat{e}_t \rightarrow RSS_U$$

<u>restricted model</u>

$$y_t = b_1 + \hat{e}_t \rightarrow RSS_R$$

F - test

$$F = \frac{(RSS_R - RSS_U)/(K-1)}{\hat{\sigma}^2_U} = \frac{ESS/(K-1)}{\hat{\sigma}^2_U}$$

under H_0 : $F \sim F_{(K-1,\, T-K)}$

1.2. TESTS OF HYPOTHESES <u>NOT</u> AVAILABLE BY DEFAULT FROM SOFTWARE

examples:

a) H_0: $\beta_1 = \beta_3 = 0$ or: $\begin{bmatrix} \beta_1 = 0 \\ \beta_3 = 0 \end{bmatrix}$

b) H_0: $\beta_2 - \beta_3 = 0$

c) H_0: $\begin{bmatrix} 50\beta_2 + 3\beta_3 = 7 \\ \beta_2 = -6 \end{bmatrix}$

F - test

$$F = \frac{(RSS_R - RSS_U)/J}{\hat{\sigma}^2_U}$$

where J is the number of statements in H_0

<u>unrestricted</u> model : $y_t = \beta_1 + \beta_2 p_t + \beta_3 a_t + e_t$

<u>restricted</u> model :

a) $y_t = \beta_2 p_t + e_t$ (gain 2 deg. of freedo

b) $y_t = \beta_1 + \beta_2 p_t - \beta_2 a_t + e_t$

$= \beta_1 + \beta_2 (p_t - a_t) + e_t$ (gain 1 deg of free

to estimate this model, you need to build a "new" regressor $(p_t - a_t)$

c) really hard to do, therefore use the approach $R\beta = r$

$$\begin{bmatrix} 0 & 50 & 3 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 7 \\ -6 \end{bmatrix}$$

<u>AND</u> use the ROLS : <u>Restricted</u> OLS

The Restricted OLS are estimators obtained from a constrained optimization of the objective function:

$$S(\beta, \lambda) = (y - X\beta)'(y - X\beta) + 2\lambda'(r - R\beta)$$

$$= \sum e_t^2 - 2\lambda'(R\beta - r)$$

where for H_0 in case c) I need

$$\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$ as I have 2 statements

$$S(\beta, \lambda) = \sum e_t^2 - 2(\lambda_1(50\beta_2 + 3\beta_3 - 7) + \lambda_2(\beta_2 + 61))$$

the solution that minimizes $S(\beta, \lambda)$ is.

b_1^*

b_2^*

b_3^*

λ_1^*

λ_2^*

2. LARGE SAMPLE

$H_0: \beta_j = 0$

under H_0:

$$\frac{b_j}{\sqrt{\hat{var}\, b_j}} \overset{asy}{\sim} N(0,1)$$

$$[N(0,1)]^2 \approx \chi^2(1)$$

$$\frac{b_j^2}{\hat{var}\, b_j} \overset{asy}{\sim} \chi^2(1)$$

if b is a vector, we have for $H_0: \beta = 0$

$$\lambda_W = b' [\hat{var}\, b]^{-1} b$$

under H_0, λ_W is asymptotically $\chi^2(J)$ distributed where J is the length of vector b or of vector β.

1. Likelihood Ratio

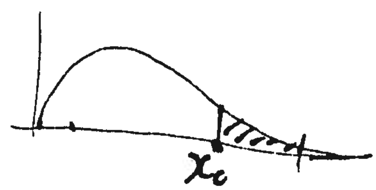
$$\lambda = \frac{\hat{L}_R}{\hat{L}_u}$$

$$H_0 : g(\theta) = q$$

$$\xi_{LR} = -2 \ln \lambda = -2 \cdot \ln \frac{\hat{L}_R}{\hat{L}_u}$$

$$= 2 \cdot \left[\ln \hat{L}_u - \ln \hat{L}_R \right]$$

under H_0, $\xi_{LR} \overset{A}{\sim} \chi^2(J)$



intuition:
when "$g(\theta) = q$" is true, $\hat{L}_u$ and $\hat{L}_R$ are close $\Rightarrow$ difference $= 0$

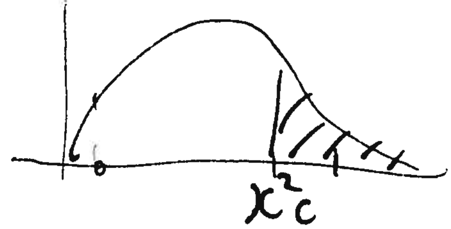
2. Wald

$$H_0 : g(\theta) = q$$

under H_0:

$$\xi_W = (g(\hat{\theta}) - q)' \left[\widehat{\text{var}}(g(\hat{\theta})) \right]^{-1} (g(\hat{\theta}) - q) \overset{A}{\sim} \chi^2(J)$$

for a scalar parameter
$H_0: \hat{\theta} = \theta_0$: $\dfrac{(\hat{\theta} - \theta_0)^2}{\widehat{\text{var}}\,\hat{\theta}} \sim \chi^2_{(1)}$



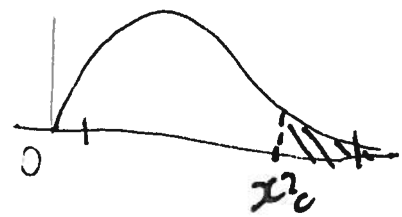
$$\text{var } g(\hat{\theta}) - q = \text{var } g(\hat{\theta}) = \frac{\partial g(\hat{\theta})}{\partial \hat{\theta}'} \left[\widehat{\text{var}}(\hat{\theta}) \right] \frac{\partial g(\hat{\theta})}{\partial \hat{\theta}}$$

DELTA METHOD

3. Lagrange Multiplier

$$\xi_{LM} = \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R}\right)' [I(\hat{\theta}_R)]^{-1} \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R}\right) \overset{A}{\sim} \chi^2(J)$$

intuition: slope of the tangent to the log-likelihood f is zero at the max for the unrestricted, and the restricted if H_0 is true



FOR ALL TESTS:

- reject H_0 if ξ_W, or ξ_{LM} or ξ_{LR} $> \chi^2(J)$

at a given $\alpha = 0.05$ for example

- accept otherwise

asymptotically, all 3 tests are equivalent

Scalar g.

CI: $\hat{\theta} \pm 1.96 \cdot \text{asy var}(\hat{\theta})$

$g(\hat{\theta}) \pm 1.96 \cdot \text{asy var}[g(\hat{\theta})]$