Inference for OLS

$$Y = X\beta + \varepsilon \qquad E\,\varepsilon_i x_i = 0 \qquad (Y_i, X_i)_{i=1}^{N} \text{ iid}$$

moments exist

Multivariate tests:

H_0: $R\beta = r$ — R is $h \times K$ matrix of linear restrictions

H_A: $R\beta \neq r$ — r is $h \times 1$

Apply Wald test

Know that $\hat{\beta} \overset{approx}{\sim} N(\beta^*, \cancel{V/N})$

where this is shorthand for

$$\sqrt{N}(\hat{\beta} - \beta^*) \xrightarrow{d} N(0, V)$$

$$V = \Sigma_{xx}^{-1} E\, x_i \varepsilon_i \varepsilon_i x_i' \underbrace{\Sigma_{xx}^{-1}}_{(E x_i x_i')}$$

Examine under H_0: $\sqrt{N}(R\hat{\beta} - r) = \sqrt{N}(R\hat{\beta} - R\beta)$

$$= \underbrace{R\sqrt{N}(\hat{\beta} - \beta)}$$

$$\xrightarrow{d} Z \sim N(0, RVR') \qquad h \text{ dimensional}$$

Take a quadratic form; use distance of $R\hat{\beta} - r$ from 0 and in particular

$$W = (R\hat{\beta} - r)'[R\hat{V}R']^{-1}(R\hat{\beta} - r)$$

where $\hat{V}$ is any consistent estimator of V.

Distribution under null is "pivotal" - does not depend on unknown parameters

So $\hat{V} \xrightarrow{p} V$ and can apply Slutsky and continuous mapping

$$N \cdot W \xrightarrow{d} z'(RVR')^{-1}z \sim \chi^2_h$$

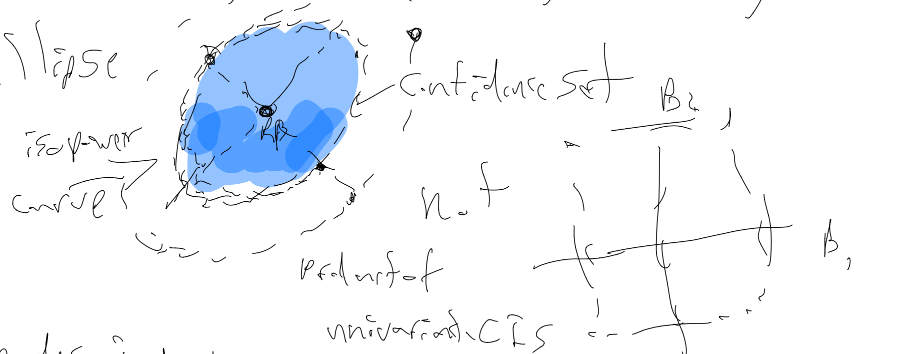
$$N \cdot W = \sqrt{N}(R\hat{B}-r)'[R\hat{V}R']^{-1}\sqrt{N}(R\hat{B}-r)$$

Reject H_0 when $N \cdot W > c_{1-\alpha}$ $\quad 1-\alpha$ quantile of χ^2_h

To get multivariate confidence set, can invert

$$CS = \{ B : N(R\hat{B}-RB)'(R\hat{V}R')^{-1}(R\hat{B}-RB) \le c_{1-\alpha} \}$$

Form is an ellipse

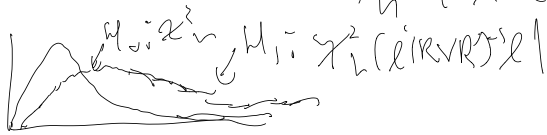


Power can be described under local alternatives

$$H_1 : RB = r + \frac{l}{\sqrt{N}}$$

Then $\sqrt{N}(R\hat{B}-r) \xrightarrow{d} N(l, V)$ under H_1

So $NW \xrightarrow{d} \chi^2_h(l'(RVR')^{-1}l)$



Power to reject is monotone in

variance weighted Euclidean distance from H_0

Easier to reject in dimensions where variance is small.

Nonlinear Hypotheses:

Want to test: $H_0: h(\theta_0) = 0$

$H_1: h(\theta_0) \neq 0$

where θ is $k \times 1$ vector arbitrary parameter with an asymptotically normal estimator $\hat{\theta}$

$h(\ )$ is an $l \times 1$ continuously differentiable vector of nonlinear restrictions on θ:

eg: $h(\theta) = \begin{pmatrix} \beta_1^2 + \beta_2^2 - 1 \\ \beta_3 \end{pmatrix}$

Assume $\underbrace{\frac{\partial h}{\partial \theta'}}_{\text{Jacobian}}$ is a full rank $l \times k$ matrix at θ_0

$\left[\frac{\partial h}{\partial \theta'}\right]_{ij}$ is $\frac{\partial h_i}{\partial \theta_j}$

We can use something like a Wald statistic

Nonlinear Wald statistic.

Assume $\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega)$ is a statistic with known asymptotic distribution (or Ω estimable)

so $\hat{\theta} \xrightarrow{p} \theta_0$

Interest is in $h(\theta)$, how to estimate?

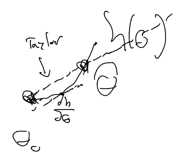
Could try CMT, but only have $\sqrt{N}(\hat{\theta} - \theta_0)$ distribution: not $\hat{\theta}$ itself.

Apply the <u>Delta Method</u>

Taylor expand $h(\hat{\theta})$ around truth θ_0

$$h(\hat{\theta}) = h(\theta_0) + \frac{\partial h}{\partial \theta'}\Big|_{\theta^+} (\hat{\theta} - \theta_0)$$

Using intermediate value form of Taylor expansion so θ^+ is between $\hat{\theta}$ and θ_0



This applies to scalar $h(\theta)$ but can apply row by row to $h(\theta)$ in vector form with possibly different θ^+ in each row of $\frac{\partial h}{\partial \theta'}$

Rearrange and normalize by $\sqrt{N}$

$$\sqrt{N}\left(h(\hat{\theta}) - h(\theta_0)\right) = \underbrace{\frac{\partial h}{\partial \theta'}\Big|_{\theta^+}}_{\xrightarrow{p} \frac{\partial h}{\partial \theta'}\big|_{\theta_0}} \underbrace{\sqrt{N}(\hat{\theta} - \theta_0)}_{\xrightarrow{d} N(0, \Omega)}$$

Note $d(\theta^+, \theta_0) \leq d(\hat{\theta}, \theta_0)$

so $\Pr\left(d(\theta^+, \theta_0) < \varepsilon\right) \geq \Pr\left(d(\hat{\theta}, \theta_0) < \varepsilon\right)$

$\rightarrow 1$ so $\theta^+ \xrightarrow{p} \theta_0$

So by continuity of derivatives $\frac{\partial h}{\partial \theta'}\Big|_{\theta^+} \xrightarrow{p} \frac{\partial h}{\partial \theta'}\Big|_{\theta_0}$

so by CMT and Slutsky

$$\sqrt{N}\left(h(\hat{\theta}) - h(\theta_0)\right) \xrightarrow{d} N\left(0, \frac{\partial h(\theta_0)}{\partial \theta'} \Omega \frac{\partial h(\theta_0)}{\partial \theta}'\right)$$

(Not same as $h\left(\sqrt{N}(\hat{\theta} - \theta_0)\right)$)

Under H_0 $h(\theta_0) = 0$, so $\sqrt{N}\left(h(\hat{\theta})\right) \xrightarrow{d} N\left(0, \frac{\partial h}{\partial \theta'} \Omega \frac{\partial h'}{\partial \theta}\right)$

Can now build a "Wald statistic"

Can rely on quadratic form convergence

$$N \cdot W = \sqrt{N}\, h(\hat{\theta})' \left(\frac{\partial h}{\partial \theta'}(\hat{\theta})\, \hat{\Omega}\, \frac{\partial h}{\partial \theta}(\hat{\theta})'\right)^{-1} \underbrace{\sqrt{N}\, h(\hat{\theta})}_{\xrightarrow{d}\ N\left(0,\ \frac{\partial h}{\partial \theta'} \Omega \frac{\partial h}{\partial \theta}'\right)}$$

$\frac{\partial h}{\partial \theta'}(\hat{\theta}) \xrightarrow{p} \frac{\partial h}{\partial \theta'}(\theta_0)$ continuity

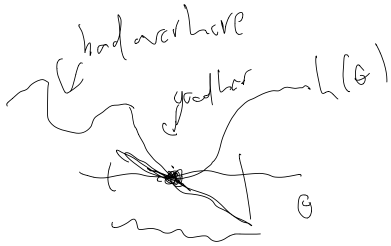
$\hat{\Omega} \xrightarrow{p} \Omega$ same consistent estimator

$\frac{\partial h}{\partial \theta}(\hat{\theta})' \xrightarrow{p} \frac{\partial h}{\partial \theta}(\theta_0)'$ continuity

So $N \cdot W \xrightarrow{d} \chi^2_l$ under H_0.

So use $c_{1-\alpha}$ from χ^2_l as critical value: reject if $NW > c_{1-\alpha}$

Delta method can fail without continuous differentiability.
Distribution may not be normal. Also needs local approximation.
So close only when sample size large enough that you are in region where Taylor expansion is accurate



by consistency, will eventually be within ε with high probability.
So can rely on this, but may need large N or small curvature

Finite sample results for OLS inference

- Quality of approximation depends on N
- Useful to know exact results in some cases
  specifically Homoskedastic Normal errors case
- Rarely if ever actually in this situation
- Good to know because reported by all standard software
- More importantly, because they give a sense of where
  asymptotic approximation loses accuracy.

(Short answer: estimated variance is not perfect;
Slutsky's theorem says you can ignore variance estimation
error if N large. In finite samples, it matters some).

Preliminaries: t and F distributions.

Let Z be standard normal and $Y \sim \chi^2_m$
independent of Z.

- Student's t distribution with m degrees of freedom
  is distribution of

$$t_m = \frac{Z}{\sqrt{Y/m}}$$

(William Sealy Gosset)

- Let $Y_1 \sim \chi^2_{m_1}$, $Y_2 \sim \chi^2_{m_2}$ be independent of each other.

- F distribution with (m_1, m_2) degrees of freedom is

$$F_{m_1, m_2} = \frac{Y_1 / m_1}{Y_2 / m_2}$$

Note that $t_m^2 = F(1, m)$ or $|t_m| = \sqrt{F(1,m)}$

because $Z^2 \sim \chi^2_1$

Note $\chi^2_m / m = \frac{1}{m} \sum_{i=1}^{m} Z_i^2 \xrightarrow{p} \underline{1}$ ← N(0,1) squared, independent

So $t_m \xrightarrow{d} Z$ as $m \to \infty$,

$F_{m_1, m_2} \xrightarrow{d} \chi^2_{m_1} / m_1$ as $m_2 \to \infty$

Classical OLS Finite Sample Results:

Let $\mathring{U} \sim N(0, \sigma^2 I_N)$ multivariate normal homoskedastic

And $Y = XB + \mathring{U}$ (OLS $\hat{B} = (X'X)^{-1} X'Y$ unbiased, has known variance)

Can use that in this case, conditional on X, $\hat{B}$ is linear transformation of normal:

$$\hat{B} = B + \underbrace{(X'X)^{-1} X' \mathring{U}}_{N(0, \sigma^2 (X'X)^{-1}) \text{ exactly}}$$

So $E[\hat{B}|X] = B$ $\quad Var(\hat{B}|X) = \sigma^2 (X'X)^{-1} = \hat{V}$

And single coefficients $\hat{\beta}_j \sim N(\beta_j, V_{jj})$

So $$\hat{z} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{V_{jj}}} \sim N(0, 1)$$

but we can't use this for testing unless we know

$$V_{jj} = [\sigma^2 (X'X)^{-1}]_{jj}$$

Replace σ^2 with $\hat{\sigma}^2$, $\underset{\uparrow}{se}(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 [(X'X)^{-1}]_{jj}}$

Standard error

Recall

$$\hat{\sigma}^2 = \frac{1}{N-K} \hat{U}'\hat{U} = \frac{1}{N-K} \mathring{U}' \underbrace{M_X}_{I - X(X'X)^{-1}X'} \mathring{U}$$

so this is $$\frac{N-K}{\sigma^2} \hat{\sigma}^2 = \underbrace{\left(\frac{\mathring{U}}{\sigma}\right)'}_{N(0,I)} \underbrace{M_X}_{\substack{\text{rank } N-K \\ \text{projection}}} \underbrace{\left(\frac{\mathring{U}}{\sigma}\right)}_{N(0,I)}$$

Apply projection result for χ^2

$$M_X = \underbrace{V}_{\text{rotation}} \underbrace{D}_{\begin{bmatrix} I_{N-K} & 0 \\ 0 & 0_K \end{bmatrix}} \underbrace{V'}_{\text{rotation}}$$

$$\hat{q} = \frac{N-K}{\sigma^2} \hat{\sigma}^2 \sim \chi^2_{N-K}$$

Also $\hat{\beta} = \beta + (X'X)^{-1} X' U$ is function of $X'\mathring{U}$

which is in span of X so covariance $X'\mathring{U}$ with

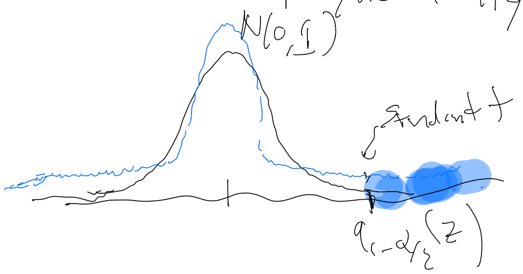
$M_X \hat{V}$ is 0 (orthogonality) so $\hat{\beta}$ and $\hat{\sigma}^2$ are functions of independent normals since orthogonality implies independence if jointly normal (not otherwise)

so
$$\frac{\hat{Z} \sim N(0,I)}{\underbrace{\sqrt{\hat{q}/(N-k)}}_{\sqrt{\chi^2_{N-k}/N-k}}} = \underbrace{\frac{\hat{\beta}_j - \beta_j}{\hat{SE}(\hat{\beta}_j)}}_{t \text{ statistic}} \sim t_{N-k}$$

exact in finite samples.

This tells us effect of error in SE estimates

t_{N-k} has fatter tails than a normal
(has only $N-k$ finite moments: tails decay polynomially rather than exponentially



$$\Pr(t > q_{1-\alpha/2}(Z)) > \Pr(N(0,1) > q_{1-\alpha/2}(Z)) = 1-\alpha/2$$

More mass in tails, and larger critical values, wider confidence intervals, especially for small α

only exactly true if errors are normal but can use it as an approximation in large N case even if not normal to account for SE estimation error; as $N \to \infty$ $t_{N-k} \to N(0,1)$ so still asymptotically valid.

Normal approximation is also not exact, so both only approximate, have the same asymptotic guarantees, but t distribution may be closer for finite N.

Multivariate Case: F-statistic.

$H_0: R\beta - r = 0$ $\quad$ R $h \times K$, r $h \times 1$

$H_1: R\beta - r \neq 0$

Test statistic is

$$\hat{F} = \frac{1}{h}(R\hat{\beta} - r)'\left(\hat{\sigma}^2 R(X'X)^{-1}R'\right)^{-1}(R\hat{\beta} - r)$$

(Homoskedastic error Wald statistic $\left(\hat{V} = \hat{\sigma}^2 (X'X)^{-1}\right)$ divided by h

Claim: under $U \sim N(0, I)$

$$\hat{F} \sim F(h, N-K) \text{ distribution}$$

Argument:

$V(R\hat{\beta}) = \sigma^2 R'(X'X)^{-1}R$

So $W = (R\hat{\beta} - r)'\left[\sigma^2 R(X'X)^{-1}R'\right]^{-1}(R\hat{\beta} - r)$

$\sim \chi^2_h$

$$\hat{F} = \frac{W/h}{\hat{\sigma}^2/N-K} \sim \frac{\chi^2_h/h}{\chi^2_{N-K}/N-K} \sim F(h, N-K)$$

independent by orthogonality

So, we have that scaled Wald statistic

has F distribution.

In large samples $N-K \to \infty$ so converges to χ^2 result but accounts for finite sample variance estimation error.

Again doesn't hurt to use F critical values (scaled by k) for Wald statistic even if not exactly normal since valid asymptotically and may be closer in finite samples.

Reported in most software: usually gives you $\hat{F}$ for test of $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ and reports p-values

Table will be of form

coef	SE	t-statistic
$\hat{\beta}_0$	$\hat{v}_{00}/\sqrt{N}$	$\hat{\beta}_0 / \hat{SE}(\hat{\beta}_0)$
$\hat{\beta}_1$	⋮	⋮
⋮		⋮
$\hat{\beta}_k$	$\hat{v}_{kk}/\sqrt{N}$	⋮

F-stat

other things (sample size, R^2)

Alternative when using robust SEs

tend "approximate" t degrees of freedom

or use something called a bootstrap

Hall "The bootstrap and edgeworth expansion"

(also a chapter in Hanson)