INTRODUCTORY ECONOMETRICS

3rd year LE & LADE

LESSON 3

Dr Javier Fernández-Macho

etpfemaj@ehu.es

Dpt. of Econometrics & Statistics

UPV—EHU

3 The Linear Regression Model (II). Inference and Prediction.

3.1a Distribution of the Least-Squares Estimator under the Normality assumption.

OLS estimator under Normality

- If $Y = X\beta + u$, where $u \sim \mathcal{N}(0, \sigma^2 I_T)$, then (recall) OLS estimator:

$$\widehat{\beta}_{\text{OLS}} = (X'X)^{-1}X'Y = \beta + (X'X)^{-1}X'u$$
$$= \beta + \Gamma' u \quad \text{is linear in disturbances.}$$

- Therefore, same Multivariate Normal distribution, with (recall)

$$\begin{cases} \text{E}(\widehat{\beta}) & = \beta, \\ \text{Var}(\widehat{\beta}) & = \sigma^2 (X'X)^{-1}. \end{cases}$$

- That is:

$$\widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2\,(X'X)^{-1})$$

OLS estimator under Normality (cases)

Since $\widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2\,(X'X)^{-1})$:

- For the k-th coefficient:

$$\widehat{\beta}_k \sim \mathcal{N}(\beta_k, \sigma^2\, a_{kk})$$

  where a_{kk} is the $(k+1)$-th diagonal element of $(X'X)^{-1}$

- *for example:* $\widehat{\beta}_1 \sim \mathcal{N}(\beta_1, \sigma^2\, a_{11})$, $a_{11} =$ 2nd diagonal element.
- For a set of linear combinations:

$$R\,\widehat{\beta} \sim \mathcal{N}(R\,\beta, \sigma^2\, R\,(X'X)^{-1}R').$$

- For a subvector of $\widehat{\beta}$: $R = [0_s \ldots 0_s | I_s]$; then

$$\widehat{\beta}^s \sim \mathcal{N}(\beta^s, \sigma^2\, A_{ss})$$

  where $\beta^s =$ subvector of β, $A_{ss} =$ submatrix of $(X'X)^{-1}$.

OLS estimator under Normality (cases)2

- In particular, if $R = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \Rightarrow$

$R \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \beta^\star$ (without intercept):

- and

$$(X'X)^{-1} = \begin{pmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{pmatrix};$$

- then

$$\widehat{\beta}^\star \sim \mathcal{N}(\beta^\star, \sigma^2 \quad)$$

OLS residuals under Normality

- Similarly, if $\boxed{u \sim \mathcal{N}(0, \sigma^2 I_T)}$,

  Then,

$$\boxed{\widehat{u} \sim \mathcal{N}(0, \sigma^2\, M)}$$

- In particular, for the 4-th residual:

$$\widehat{u}_t \sim \mathcal{N}(0, \sigma^2\, m_{44})$$

  where m_{44} is the 4-th diagonal element of matrix M.

3.1b Hypothesis Testing: a Review.

Hypothesis and Tests (rev1)

- Starting point:

$$\left.\begin{array}{ll} Y & = X\beta + u \\ u & \sim \mathcal{N}(0, \sigma^2 I_T) \end{array}\right\} \left\{\begin{array}{l} \widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (X'X)^{-1}) \\ \widehat{u} \sim \mathcal{N}(0, \sigma^2 M) \end{array}\right.$$

- Hypothesis: "conjecture about parameter(s) dn fn".
  For example:
  - in SLRM: $\widehat{\beta} \sim \mathcal{N}(\beta, v)$; assume $\beta = 2.5$.
  - in GLRM: $\widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (X'X)^{-1})$; assume $\beta_1 + \cdots + \beta_K = 1$.
  - in general: Ec. Th. $\rightsquigarrow$ hypothesis
    e.g.: Cobb-Couglas Fn:

$$Y_t = e^{\beta_0} L_t^{\beta_1} K_t^{\beta_2} e^{u_t}$$

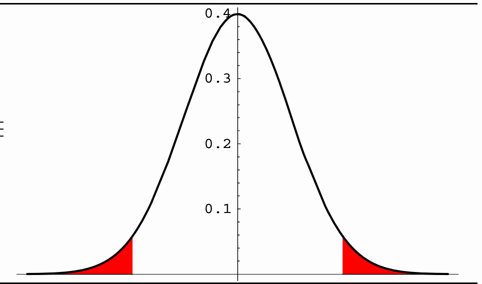
    with Constant returns to scale: $\beta_1 + \beta_2 = 1$
- Test: "procedure to reject or accept the hypothesis"

Hypothesis and Tests (rev2)

	elements	steps
a)	hypothesis to test (about estimator)	$H_0 : \ldots$ vs. $H_a : \ldots$ (disjoint)
b)	estimator dn	obtain test statistic with tabulated dn under H_0 :
c)	decision rule	calculated statistic: $\in$ critical region ("large") $\Downarrow$ Reject; $\notin$ critical region ("small") $\Downarrow$ not Reject

Hypothesis and Tests (rev2-cont)

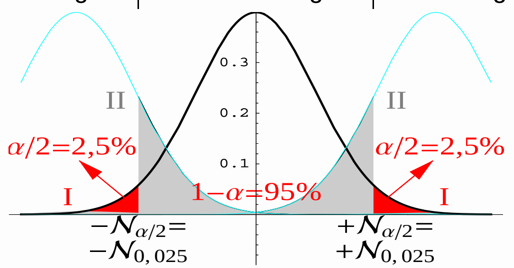
Example:

a)	$H_0 : \beta = 2.5$ vs. $H_a : \beta \neq 2.5$	(Var(β)=4)
b)	$\widehat{\beta} \sim \mathcal{N}(\beta, 4) \rightsquigarrow z = \dfrac{\widehat{\beta} - \beta}{2} \sim \mathcal{N}(0,1)$	
c)	$z = \dfrac{\widehat{\beta} - 2.5}{2} \in$ 	

Hypothesis and Tests: Critical region

significance level $\alpha = 5\% = 0.05$

critical region	confidence region	critical region

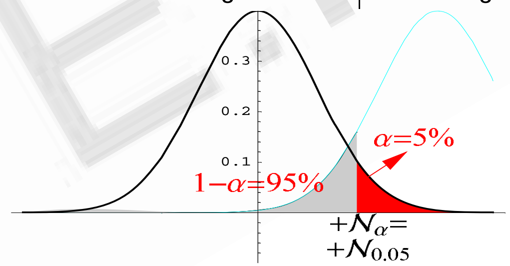


critical region	confidence region	critical region

Hypothesis and Tests: Critical region (one sided)

significance level $\alpha = 5\% = 0.05$

confidence region	critical region



confidence region	critical region

Hypothesis and Tests: Distributions (rev)

1. Def of χ^2 (chi-square):

$$\left.\begin{array}{r} Z_i \sim \text{iid}\, \mathcal{N}(0,1) \\ Z \sim \mathcal{N}(0, I_m) \end{array}\right\} Z'Z = \sum_{i=1}^{m} Z_i^2 \sim \chi^2(m) \qquad \left\{\begin{array}{l} \mathrm{E}(\chi^2(m)) = m \\ \mathrm{Var}(\chi^2(m)) = 2m \end{array}\right.$$

1b. $Z \sim \mathcal{N}(\mu, \Omega) \Rightarrow (Z-\mu)'\Omega^{-1}(Z-\mu) \sim \chi^2(m)$

2. Def of t (Student):

$$\left.\begin{array}{r} Z \sim \mathcal{N}(0,1), \quad W \sim \chi^2(m) \\ Z, W \text{ independent} \end{array}\right\} \frac{Z}{\sqrt{W/m}} \sim t(m)$$

3. Def of $\mathcal{F}$ (Snedecor):

$$\left.\begin{array}{r} V \sim \chi^2(n) \quad W \sim \chi^2(m) \\ V, W \text{ independent} \end{array}\right\} \frac{V/n}{W/m} \sim \mathcal{F}^n_m$$

4b. $n = 1 \Rightarrow \dfrac{Z^2}{W/m} \sim \mathcal{F}^1_m \equiv t(m)^2$

Hypothesis and Tests: Useful result

From $\widehat{u} \sim \mathcal{N}(0, \sigma^2 M)$:

- $\dfrac{\text{RSS}}{\sigma^2} = \sum(\widehat{u}_t^2/\sigma^2) = \sum \mathcal{N}(0,1)^2\text{'s} \sim \chi^2(T–K–1)$
- Then: $\dfrac{\widehat{\sigma}^2}{\sigma^2} = \dfrac{\text{RSS}}{\sigma^2(T–K–1)} = \dfrac{\text{RSS}}{\sigma^2(T–K–1)} = \chi^2/\text{d.f.'s}$
  - $\dfrac{\text{expr}}{\sigma} \sim \mathcal{N}(0,1)$:
  - $\dfrac{\text{expr}}{\widehat{\sigma}} = \dfrac{\text{expr}/\sigma}{\widehat{\sigma}/\sigma} = \dfrac{\text{expr}/\sigma}{\sqrt{\widehat{\sigma}^2/\sigma^2}} = \dfrac{\mathcal{N}(0,1)}{\sqrt{\chi^2/\text{d.f.'s}}} = t$
  - $\dfrac{\text{expr}}{\sigma^2} \sim \chi^2(n)$:
  - $\dfrac{\text{expr}}{\widehat{\sigma}^2} = \dfrac{\text{expr}/\sigma^2}{\widehat{\sigma}^2/\sigma^2} \Rightarrow \dfrac{\frac{\text{expr}}{\sigma^2}/n}{\widehat{\sigma}^2/\sigma^2} = \dfrac{\chi^2(n)/n}{\chi^2/\text{d.f.'s}} \sim \mathcal{F}$
- In short: $\sigma^2 \to \widehat{\sigma}^2 \Rightarrow$ $\mathcal{N}(0,1) \to t$!! $\chi^2 \to \mathcal{F}$!!

3.2a Testing for the Significance of a single parameter. Confidence Intervals.

Single parameter Significance test: estimator dn

- Standardise $\widehat{\beta}_i \sim \mathcal{N}(\beta_i, \sigma^2 a_{ii})$

$$\frac{\widehat{\beta}_i - \beta_i}{\sqrt{\mathrm{Var}(\widehat{\beta}_i)}} = \frac{\widehat{\beta}_i - \beta_i}{\sigma\sqrt{a_{ii}}} = \frac{\widehat{\beta}_i - \beta_i}{\sigma_{\widehat{\beta}_i}} \sim \mathcal{N}(0,1)$$

- change σ by $\widehat{\sigma}$:

$$\frac{\widehat{\beta}_i - \beta_i}{\widehat{\sigma}\sqrt{a_{ii}}} = \frac{\widehat{\beta}_i - \beta_i}{\sqrt{\widehat{\mathrm{Var}(\widehat{\beta}_i)}}} = \boxed{\frac{\widehat{\beta}_i - \beta_i}{S_{\widehat{\beta}i}} \sim t(T–K–1)}$$

- Note how $\sigma_{\widehat{\beta}i} \to S_{\widehat{\beta}i} \Rightarrow \mathcal{N}(0,1) \to t$!!

Single parameter Significance test: rule

- $\dfrac{\widehat{\beta}_i - \beta_i}{S_{\widehat{\beta}i}} \sim t_{(T-K-1)}$
- Which Test? $\begin{cases} H_0 : \beta_i = c & \text{(informative test)} \\ H_0 : \beta_i = 0 & \text{(test of significance)} \end{cases}$
- **Remember:** Hypothesis $\rightsquigarrow$ statistic $\rightsquigarrow$ rule...
- Test of Significance:
  - Hypothesis: $H_0 : \beta_i = 0$ vs. $H_a : \beta_i \neq 0$
  - Statistic: $t = \dfrac{\widehat{\beta}_i}{S_{\widehat{\beta}i}} \sim t_{(T-K-1)}$ under H_0 :
  - Rule: $|t| = |\dfrac{\widehat{\beta}_i}{S_{\widehat{\beta}i}}| > t_{\alpha/2}(T-K-1) \quad \Rightarrow$ reject H_0 :

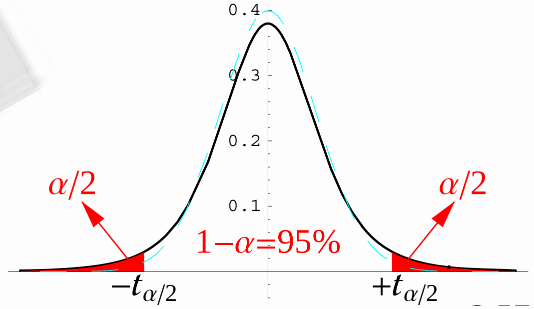
    $\Rightarrow$ β_i is (statistically or significantly) different from zero

    $\Rightarrow$ X_i is a (statistically) relevant or significant variable.
- similarly for informative test $H_0 : \beta_i = c$ (Exercise: Try it!!)

Single parameter Significance test: rule (cont)

- Rule: $|t| = |\dfrac{\widehat{\beta}_i}{S_{\widehat{\beta}i}}| > t_{\alpha/2}(T-K-1) \quad \Rightarrow$ reject H_0 :
- significance level $\alpha = 5\% = 0.05$

critical region	confidence region	critical region

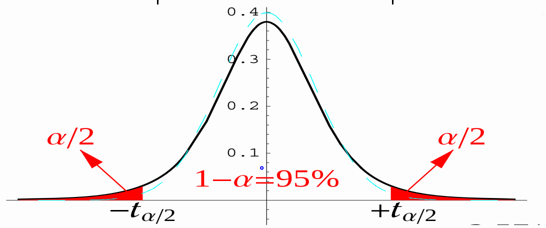


critical region	confidence region	critical region

Confidence interval for β_i

- Recall that $\dfrac{\widehat{\beta}_i - \beta_i}{S_{\widehat{\beta}i}} \sim t_{(T-K-1)}$

critical region	confidence region	critical region



critical region	confidence region	critical region

- *i.e.:* $\Pr[-t_{\alpha/2} \leq \dfrac{\widehat{\beta}_i - \beta_i}{S_{\widehat{\beta}i}} \leq +t_{\alpha/2}] = 1 - \alpha$

$$\Pr[\underbrace{\widehat{\beta}_i - t_{\alpha/2}\, S_{\widehat{\beta}i} \leq \beta_i \leq \widehat{\beta}_i + t_{\alpha/2}\, S_{\widehat{\beta}i}}_{\text{CI}_{1-\alpha}(\beta_i)}] = 1 - \alpha$$

Confidence interval for β_i (cont)

- That is:

$$\text{CI}_{1-\alpha}(\beta_i) = [\widehat{\beta}_i \pm t_{\alpha/2}\, S_{\widehat{\beta}i}]$$

- *e.g.* for $\alpha = 5\%$, $T-K-1 = 25$, $\widehat{\beta}_i = 2.12$ and $S_{\widehat{\beta}i} = 0.08$:

$$\text{CI}_{95\%}(\beta_i) = [\widehat{\beta}_i \pm t_{2.5\%}(25)\; S_{\widehat{\beta}i}]$$
$$= [\widehat{\beta}_i \pm 2.06\; S_{\widehat{\beta}i}] = [2.12 \pm 2.06 \cdot 0.08] = [1.9552\,;\,2.2848]$$

- testing by means of confidence interval:

Hypothesis:	$H_0 : \beta_i = c$ vs. $H_a : \beta_i \neq c$
Interval:	$\text{CI}_{95\%}(\beta_i)$
Rule:	Reject H_0 : if $c \notin \text{CI}_{95\%}(\beta_i)$, with 5% significance.

- *e.g.* $H_0 : \beta_i = 0$? $\Rightarrow$ Reject $\Rightarrow$ β_i is significant (at 5% level).

Testing a Single Linear Combination

- Let's have a restricted GLRM with 1 restriction ($q = 1$):
  $R\beta = r$ but now simpler...
  $R = d'$ (any row of $K{+}1$ values $d_0, d_1, \ldots, +d_K$) and
  $r = c$ (any single value):
- Let $H_0 : \nu = d'\beta = d_0\beta_0 + d_1\beta_1 + \cdots + d_K\beta_K = c$
  that is,
  an informative test about the value c that takes a single linear combination ν of the parameters.

Testing a Single Linear Combination: Example

- Let's have the linearised Cobb-Douglas fn

$$\log Y_t = \alpha + \beta_L \log L_t + \beta_K \log K_t + u_t$$

$d' = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}$ and $c = 1$:

- $$H_0 : \nu = d'\beta = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \begin{pmatrix} \alpha \\ \beta_L \\ \beta_K \end{pmatrix} \qquad = \beta_L + \beta_K$$
$$= c \qquad\qquad = 1$$

  that is, $H_0 : \beta_L + \beta_K = 1$;
  the test of the constant returns to scale hypothesis.

Testing a Single Linear Combination: dn

- Since $\quad \widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X'X)^{-1}), \quad$ we have that

$$d'\widehat{\beta} \sim \mathcal{N}(d'\beta, \sigma^2 d'(X'X)^{-1}d)$$
$$\widehat{\nu} \sim \mathcal{N}(\nu, \text{Var}(\widehat{\nu}))$$

  where $\text{Var}(\widehat{\nu}) = \sigma^2 \sum_{i,j=0}^{K} d_i d_j a_{ij}$
- As before, standardise $\widehat{\nu}$

$$\frac{\widehat{\nu} - \nu}{\sqrt{\text{Var}(\widehat{\nu})}} \sim \mathcal{N}(0,1)$$

- Therefore (recall $\sigma \to \widehat{\sigma}$):

$$\Rightarrow \boxed{\frac{\widehat{\nu} - \nu}{S_{\widehat{\nu}}} \sim t_{(T-K-1)}}$$

  where $S_{\widehat{\nu}} = \widehat{\sigma}\sqrt{\sum_{i,j=0}^{K} d_i d_j a_{ij}}$.

Testing a Single Linear Combination: rule

- $\boxed{\frac{\widehat{\nu} - \nu}{S_{\widehat{\nu}}} \sim t_{(T-K-1)}}$
- Which Test? $\{ H_0 : \nu (= d'\beta) = c$ (informative test)
- **Remember:** Hypothesis $\rightsquigarrow$ statistic $\rightsquigarrow$ rule...
- Test for a linear combination:
  - Hypothesis: $H_0 : \nu = c$ vs. $H_a : \nu \neq c$
  - Statistic:

$$t = \frac{\widehat{\nu} - c}{S_{\widehat{\nu}}} \sim t_{(T-K-1)} \text{ under } H_0 :$$

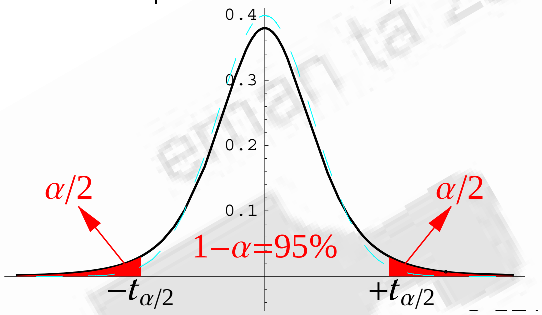
  - Rule: $\quad |t| > t_{\alpha/2(T-K-1)} \quad \Rightarrow \quad$ reject H_0 :
    $\Rightarrow$ value of linear combination isn't right.
  - *cf* test of single parameter β_k, any similarities?.

Testing a Single Linear Combination: rule (cont)

- Rule: $|t| = \left| \frac{\widehat{\nu} - c}{S_{\widehat{\nu}}} \right| > t_{\alpha/2}(T\text{–}K\text{–}1) \quad \Rightarrow \quad$ reject H_0 :
-

significance level $\alpha = 5\% = 0.05$

critical region	confidence region	critical region



critical region	confidence region	critical region

Testing a Single Linear Combination: Example

- In the linearised Cobb-Douglas fn:
  $\widehat{\log Y}_t = \widehat{\alpha} + \widehat{\beta}_L \log L_t + \widehat{\beta}_K \log K_t, \quad T = 53;$
- $\widehat{\log Y}_t = 2.10 + 0.67 \log L_t + 0.32 \log K_t, \quad \widehat{\sigma}^2 = 4$;

$$(X'X)^{-1} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & -1 \\ 0 & -1 & 7 \end{pmatrix}$$

- Test the H_0 : constant returns to scale
  at the $\alpha = 5\%$ significance level:

Testing a Single Linear Combination: Example

- Hypothesis: $H_0 : \beta_L + \beta_K = 1$ vs. $H_a : \beta_L + \beta_K \neq 1$
- Statistic:

$$\widehat{\nu} = \widehat{\beta}_L + \widehat{\beta}_K$$
$$= 0.67 + 0.27 = 0.89$$

◆

$$S_{\widehat{\nu}} = \sqrt{\widehat{\text{Var}}(\widehat{\beta}_L) + \widehat{\text{Var}}(\widehat{\beta}_K) + 2\widehat{\text{Cov}}(\widehat{\beta}_L, \widehat{\beta}_K)}$$
$$= \widehat{\sigma}\sqrt{a_{11} + a_{22} + 2a_{12}}$$
$$= 2\sqrt{4 + 7 + 2(-1)} = 2\sqrt{9} = 6$$

◆

$$t = \frac{\widehat{\nu} - 1}{S_{\widehat{\nu}}}$$
$$= \frac{0.89 - 1}{6} = \frac{-0.11}{6} = -0.018.$$

- Rule: $|t| = 0.018 < t_{0.025}(50) = 2.01 \quad \Rightarrow$ don't reject H_0 :
  $\Rightarrow$ "constant returns to scale" is supported by data.

3.2b Testing for Overall Significance.

Overall Significance Test: estimator dn

- $H_0 : \beta_1 = \beta_2 = \cdots = \beta_K = 0 \quad \rightsquigarrow$
- $H_0 : \beta^\star = \mathbf{0} \quad \rightsquigarrow$
- $$\widehat{\beta}^\star \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \begin{bmatrix} a_{00} & a_{01} & \dots & a_{0K} \\ a_{10} & a_{11} & \dots & a_{1K} \\ \vdots & \vdots & \dots & \vdots \\ a_{K0} & a_{K1} & \dots & a_{KK} \end{bmatrix}\right) \sim \mathcal{N}(\mathbf{0}, \sigma^2 (x'x)^{-1})$$
- Standardise and SS:
  - $$\frac{\widehat{\beta}^{\star\prime} x' x \widehat{\beta}^\star}{\sigma^2} \sim \chi^2(K) \quad \text{under } H_0 :$$
- Therefore (recall changing $\sigma^2 \to \widehat{\sigma}^2$):

$$F = \frac{\widehat{\beta}^{\star\prime} x' x \widehat{\beta}^\star / K}{\widehat{\sigma}^2} \sim \mathcal{F}^K_{T\text{-}K\text{-}1}$$

Overall Significance Test: rule

- $$F = \frac{\widehat{\beta}^{\star\prime} x' x \widehat{\beta}^\star / K}{\widehat{\sigma}^2} \sim \mathcal{F}^K_{T\text{-}K\text{-}1} \quad \text{under } H_0 :$$
- Overall significance test: $\{H_0 : \beta^\star = 0$
- **Remember:** Hypothesis $\rightsquigarrow$ statistic $\rightsquigarrow$ rule...
  - Hypothesis: $H_0 : \beta^\star = 0$ vs. $H_a : \beta^\star \neq 0$ (*i.e.* $\exists \beta_i \neq 0$)
  - Statistic:

$$F = \frac{\widehat{\beta}^{\star\prime}\, x' x\, \widehat{\beta}^\star / K}{\widehat{\sigma}^2} = \frac{\widehat{y}'\widehat{y} / K}{\widehat{u}'\widehat{u} / (T\text{–}K\text{–}1)} = \frac{\text{ESS} / K}{\text{RSS} / (T\text{–}K\text{–}1)}$$
$$= \frac{(\text{ESS}/\text{TSS}) / K}{(\text{RSS}/\text{TSS}) / (T\text{–}K\text{–}1)} = \frac{R^2 / K}{(1 - R^2) / (T\text{–}K\text{–}1)} \sim \mathcal{F}^K_{T\text{-}K\text{-}1} \text{ under } H_0 :$$

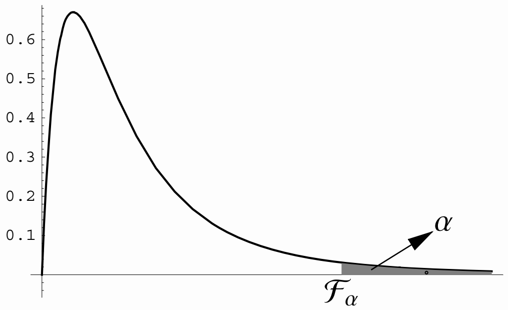
  - Rule: $F > \mathcal{F}_\alpha(K, T\text{–}K\text{–}1) \quad \Rightarrow$ reject H_0 :
    - $\Rightarrow$ all coefs are jointly significant (different from zero)
    - $\Rightarrow$ whole regression is (statistically) relevant.

Overall Significance Test: rule (cont)

- Rule: $F > \mathcal{F}_\alpha(K, T\text{–}K\text{–}1) \quad \Rightarrow \quad$ reject H_0 :
-

significance level $\alpha = 5\% = 0.05$

confidence region	critical region



confidence region	critical region

Overall Significance Test: Example

- In the previous example (linearised Cobb-Douglas fn:)

$$\widehat{\log Y}_t = \widehat{\alpha} + \widehat{\beta}_L \log L_t + \widehat{\beta}_K \log K_t, \quad T = 53;$$
$$\widehat{\log Y}_t = 2.10 + 0.67 \log L_t + 0.32 \log K_t, \quad \widehat{\sigma}^2 = 4; R^2 = 0.88$$

- Test the overall significance at the $\alpha = 5\%$ significance level:
-

$$F = \frac{R^2 / K}{(1 - R^2) / (T\text{–}K\text{–}1)}$$
$$= \frac{0.88 / 2}{(1 - 0.88) / (50)} = \frac{0.44}{0.024} = 183.33 > \mathcal{F}_{0.05}(2, 50) = 3.19$$

$\Rightarrow$ β_K & β_L are jointly significant
$\Rightarrow$ regression is (statistically) relevant.

3.3 A General Test for Linear Restrictions.

Testing for Linear Restrictions: Example 1

- Recall GLRM subject to q linear restrictions:

$$\underset{(T\times 1)}{Y} = \underset{(T\times K{+}1)}{X}\ \underset{(K{+}1\times 1)}{\beta} + \underset{(T\times 1)}{u},$$
$$H_0: \underset{(q\times K{+}1)}{R}\ \underset{(K{+}1\times 1)}{\beta} = \underset{(q\times 1)}{r}.$$

- Previous tests $\equiv$ special cases LRs:
  1. Let's have the GLRM with $q=1$, $R = \begin{bmatrix} 0 & 0 & 1 & \dots & 0 \end{bmatrix}$ and $r=0$:

$$H_0: R\beta = \begin{bmatrix} 0 & 0 & 1 & \dots & 0 \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_K \end{pmatrix} = \beta_2$$
$$= r \qquad\qquad = 0$$

i.e., $H_0: \beta_2 = 0$;

the test of individual significance of X_2.

Testing for Linear Restrictions: Example 2

- $H_0: \underset{(q\times K{+}1)}{R}\ \underset{(K{+}1\times 1)}{\beta} = \underset{(q\times 1)}{r}.$
  2. Let's assume $q = K$ restrictions such that

$$R = \begin{bmatrix} \mathbf{0} & | & \mathbf{I}_K \end{bmatrix} = \left(\begin{array}{c|cccc} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{array}\right) \text{ and } r = \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- $$H_0: R\beta = \begin{bmatrix} \mathbf{0}_K & | & \mathbf{I}_K \end{bmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta^\star \end{pmatrix} = \beta^\star$$
$$= r \qquad\qquad = \mathbf{0}$$

that is, $H_0: \beta^\star = \mathbf{0}$;

the test of overall significance of the regression.

Testing for Linear Restrictions: Example 3

- $H_0: \underset{(q\times K{+}1)}{R}\ \underset{(K{+}1\times 1)}{\beta} = \underset{(q\times 1)}{r}.$
  3. Let's assume $q = 2$ restrictions such that

$$R = \begin{bmatrix} 0 & 2 & 3 & 0 & \dots & 0 \\ 1 & 0 & 0 & -2 & \dots & 0 \end{bmatrix} \text{ and } r = \begin{bmatrix} 5 \\ 3 \end{bmatrix}:$$

- $$H_0: R\beta = \begin{bmatrix} 0 & 2 & 3 & 0 & \dots & 0 \\ 1 & 0 & 0 & -2 & \dots & 0 \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \dots \\ \beta_K \end{pmatrix} = \begin{bmatrix} 2\beta_1 + 3\beta_2 \\ \beta_0 - 2\beta_3 \end{bmatrix}$$
$$= r \qquad\qquad = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

that is, the GLRM under $H_0: \begin{cases} 2\beta_1 + 3\beta_2 = 5 \\ \beta_0 - 2\beta_3 = 3 \end{cases}$

Testing for Linear Restrictions: dn

- ...so, can have a general test statistic to cover for all hypothesis of the form

$$H_0: \underset{(q\times K+1)}{R} \underset{(K+1\times 1)}{\beta} = \underset{(q\times 1)}{r} \quad ?$$

- Given that $\widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (X'X)^{-1})$, we have that

$$R\widehat{\beta} \sim \mathcal{N}(R\beta, \sigma^2 R(X'X)^{-1}R')$$

- As before, standardise $R\widehat{\beta}$ and construct SS,

$$\frac{(R\widehat{\beta} - R\beta)'[R(X'X)^{-1}R']^{-1}(R\widehat{\beta} - R\beta)}{\sigma^2} \sim \chi^2(q)$$

- Therefore (recall changing $\sigma^2 \to \widehat{\sigma}^2$):

$$\boxed{\frac{(R\widehat{\beta} - R\beta)'[R(X'X)^{-1}R']^{-1}(R\widehat{\beta} - R\beta)/q}{\widehat{\sigma}^2} \sim \mathcal{F}^q_{T\text{-}K\text{-}1}}$$

General Test for Linear Restrictions: rule

- Which Test? $\{H_0 : R\beta = r$
- **Remember:** Hypothesis ⇝ statistic ⇝ rule...
- Test for linear restrictions:
  - Hypothesis: $H_0 : R\beta = r$ vs. $H_a : R\beta \neq r$
  - Statistic:

$$F = \frac{(R\widehat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\widehat{\beta} - r)/q}{\widehat{\sigma}^2} \sim \mathcal{F}^q_{T\text{-}K\text{-}1} \text{ under } H_0 :$$

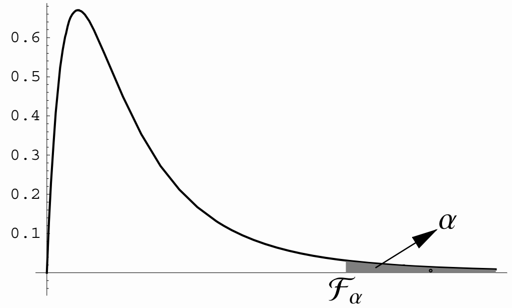
  - Rule: $F > \mathcal{F}_\alpha(q, T\text{–}K\text{–}1) \Rightarrow$ reject H_0 :
    $\Rightarrow$ linear restrictions aren't (jointly) true.

General Test for Linear Restrictions: rule (cont)

- Rule: $F > \mathcal{F}_\alpha(q, T\text{–}K\text{–}1) \Rightarrow$ reject H_0 :
-

significance level $\alpha = 5\% = 0.05$

confidence region | critical region



confidence region | critical region

3.4 Tests based on the Residual Sum of Squares.

General Test for Linear Restrictions: rule 2

- Hypothesis: $H_0 : R\beta = r$ vs. $H_a : R\beta \neq r$
- Statistic:

$$F = \frac{(R\widehat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\widehat{\beta} - r)/q}{\widehat{\sigma}^2}$$

- using result on $\widehat{\beta}_R = (I - AR)\widehat{\beta} + Ar$, numerator is difference between SS's:

$$F = \frac{(\text{RSS}_R - \text{RSS})/q}{\text{RSS}/(T\text{–}K\text{–}1)} \sim \mathcal{F}^q_{T\text{-}K\text{-}1} \text{ under } H_0 :$$

- Rule: $F > \mathcal{F}_\alpha(q, T\text{–}K\text{–}1) \Rightarrow$ reject H_0 :
  $\Rightarrow$ linear restrictions aren't (jointly) true.

General Test for Linear Restrictions: Summary

- Hypothesis: $H_0 : R\beta = r$ vs. $H_a : R\beta \neq r$
- Statistic:

$$F = \frac{(R\widehat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\widehat{\beta} - r)/q}{\widehat{\sigma}^2}$$

$$= \frac{(\text{RSS}_R - \text{RSS})/q}{\text{RSS}/(T\text{–}K\text{–}1)} \sim \mathcal{F}^q_{T\text{-}K\text{-}1} \text{ under } H_0 :$$

- Rule: $F > \mathcal{F}_\alpha(q, T\text{–}K\text{–}1) \Rightarrow$ reject H_0 :
  $\Rightarrow$ linear restrictions aren't (jointly) true.
- Note that, SS form needs estimating twice: unrestricted and restricted regressions.
- and, of course, they can also be used to test for individual significance, overall significance, informative restrictions, etc.

Test based on SS: Example Cobb-Douglas

- Hypothesis: $H_0 : \beta_L + \beta_K = 1$ vs. $H_a : \beta_L + \beta_K \neq 1$
- Statistic:
  -
$$\widehat{\nu} = \widehat{\beta}_L + \widehat{\beta}_K = 0.67 + 0.27 = 0.89$$
  -
$$S_{\widehat{\nu}} = \sqrt{\widehat{\text{Var}}(\widehat{\beta}_L) + \widehat{\text{Var}}(\widehat{\beta}_K) + 2\widehat{\text{Cov}}(\widehat{\beta}_L, \widehat{\beta}_K)} = \widehat{\sigma}\sqrt{a_{11} + a_{22} + 2a_{12}} = 2\sqrt{4 + 7 + 2(-1)} = 2\sqrt{9} = 6$$
  -
$$t = \frac{\widehat{\nu} - 1}{S_{\widehat{\nu}}} = \frac{0.89 - 1}{6} = \frac{-0.11}{6} = -0.018.$$
- Rule: $|t| = 0.018 < t_{0.025}(50) = 2.01 \Rightarrow$ don't reject H_0 :
  $\Rightarrow$ "constant returns to scale" is supported by data.

Test based on SS: Example Cobb-Douglas

- Alternatively, use SS form to calculate this t ratio:
  unrestricted:
  $\log Y = \alpha + \beta_L \log L + \beta_K \log K + u, \quad \rightsquigarrow \quad \text{RSS} = 200$
- restricted: $\log Y = \alpha + \beta_L \log L + (1 - \beta_L) \log K + u$
  $\log(Y/K) = \alpha + \beta_L \log(L/K) + u, \quad \rightsquigarrow \quad \text{RSS}_R = 200.001296$
-
$$F = \frac{(\text{RSS}_R - \text{RSS})/q}{\text{RSS}/(T\text{–}K\text{–}1)} = \frac{(200.001296 - 200)/1}{200/50} = \frac{.001296}{4} = 0.000324 < \mathcal{F}_{0.05}(1, 50) = 4.04$$
- or (recall $t(m) = \sqrt{\mathcal{F}(1, m)}$)

$$t = \sqrt{F} = \sqrt{0.000324} = 0.018 < t_{0.05}(50) = 2.01$$

General Test: Example 3

- GLRM with $q = 2$, $R = \begin{bmatrix} 0 & 2 & 3 & 0 & \dots & 0 \\ 1 & 0 & 0 & -2 & \dots & 0 \end{bmatrix}$ and $r = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$:

$$R\widehat{\beta} = \begin{bmatrix} d_1'\widehat{\beta} \\ d_2'\widehat{\beta} \end{bmatrix} = \begin{bmatrix} 2\widehat{\beta}_1 + 3\widehat{\beta}_2 \\ \widehat{\beta}_0 - 2\widehat{\beta}_3 \end{bmatrix}$$

-

$$R(X'X)^{-1}R' = \begin{bmatrix} d_1'(X'X)^{-1}d_1 & d_1'(X'X)^{-1}d_2 \\ d_2'(X'X)^{-1}d_1 & d_2'(X'X)^{-1}d_2 \end{bmatrix}$$
$$= \begin{bmatrix} 4a_{11} + 9a_{22} + 6a_{12} & 2a_{10} - 4a_{13} + 3a_{20} - 6a_{23} \\ & a_{00} + 4a_{33} - 2a_{03} \end{bmatrix}$$

- Therefore $F =$

$$\frac{\begin{bmatrix} 2\beta_1 + 3\beta_2 - 5 & \beta_0 - 2\beta_3 - 3 \end{bmatrix} \begin{bmatrix} 4a_{11} + 9a_{22} + 6a_{12} & 2a_{10} - 4a_{13} + 3a_{20} - 6a_{23} \\ & a_{00} + 4a_{33} - 2a_{03} \end{bmatrix}^{-1} \begin{bmatrix} 2\beta_1 + 3\beta_2 - 5 \\ \beta_0 - 2\beta_3 - 3 \end{bmatrix} /2}{\widehat{\sigma}^2}$$

$\sim \mathcal{F}^2_{T-K-1}$ under H_0 :

- *i.e.*, an "F" statistic for testing two linear restrictions jointly.

General Test: Example 3

- Alternatively (easier), use SS form to calculate this F statistic:

$$H_0 : \begin{cases} 2\beta_1 + 3\beta_2 = 5 \\ \beta_0 - 2\beta_3 = 3 \end{cases}$$
$$\beta_1 = \frac{5 - 3\beta_2}{2}, \qquad \beta_0 = 3 + 2\beta_3$$

- unrestricted:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \cdots + u \rightsquigarrow \text{RSS}$$

- restricted:

$$Y = (3 + 2\beta_3) + (2.5 - 1.5\beta_2) X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \cdots + u$$
$$\underbrace{Y - 3 - 2.5X_1}_{Y^*} = \beta_2 \underbrace{(X_2 - 1.5X_1)}_{X_2^*} + \beta_3 \underbrace{(X_3 + 2)}_{X_3^*} + \beta_4 X_4 \cdots + u$$
$$Y^* = \beta_2 X_2^* + \beta_3 X_3^* + \beta_4 X_4 \cdots + u \rightsquigarrow \text{RSS}_R$$

- and $F = \frac{(\text{RSS}_R - \text{RSS})/q}{\text{RSS}/(T-K-1)}$, etc.

General Test: Example 2

- GLRM with $q = K$, $R = \begin{bmatrix} \mathbf{0}_K & | & \mathbf{I}_K \end{bmatrix}$ and $r = \mathbf{0}_K$:

$$R\widehat{\beta} \quad \rightsquigarrow \text{selects} \quad \beta^\star$$

$$R(X'X)^{-1}R' \quad \rightsquigarrow \text{selects} \quad \begin{bmatrix} a_{00} & a_{01} & \dots & a_{0K} \\ a_{10} & a_{11} & \dots & a_{1K} \\ \vdots & \vdots & \dots & \vdots \\ a_{K0} & a_{K1} & \dots & a_{KK} \end{bmatrix} = (x'x)^{-1}$$

- Therefore:

$$F = \frac{(\widehat{\beta}^\star - 0)'[(x'x)^{-1}]^{-1}(\widehat{\beta}^\star - 0)/K}{\widehat{\sigma}^2}$$
$$= \frac{\widehat{\beta}^{\star\prime}\, x'x\, \widehat{\beta}^\star / K}{\widehat{\sigma}^2}$$

- *i.e.*, the usual "F" statistic for testing the overall significance of the regression.

General Test: Example 2

- Alternatively, use SS form to calculate this F:

unrestricted: $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K + u \quad \rightsquigarrow \quad \text{RSS}$

restricted: $Y = \beta_0 + u \quad \rightsquigarrow \quad \text{RSS}_R = \text{TSS}$

- Statistic:

$$F = \frac{(\text{RSS}_R - \text{RSS})/q}{\text{RSS}/(T-K-1)} = \frac{(\text{TSS} - \text{RSS})/K}{\text{RSS}/(T-K-1)} = \frac{\text{ESS}/K}{\text{RSS}/(T-K-1)}$$
$$= \frac{R^2/K}{(1 - R^2)/(T-K-1)}$$

obtaining same formula as before.

3.5 Point Prediction and Prediction Interval.

Prediction

- Previous chapters: **Specification, Estimate and Validation**.
- This chapter: Final stage: **Use = Prediction**.
- Starting point: appropriate model to describe behaviour of variable Y:

Model Specification
⇓
Parameter Estimation
⇓
Validation and Hypothesis Testing

- ⇓

not satisfactory result	satisfactory result
⇓	⇓
respecify model	go to prediction stage

Concept

- Time series: prediction (of future values)
  ⇒ Forecasting
- Cross-section: prediction (of unobserved values)
  ⇒ Simulation
- In general: prediction ⇒ answer to "what if. . . ?" questions,
  i.e. what value would take Y if $X = X_p$?

Basic Elements

- Model or PRF:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_K X_{Kt} + u_t$$
$$Y_t = X_t'\beta + u_t, \quad t = 1, \ldots, T.$$

- Estimated model or SRF:

$$\widehat{Y}_t = X_t'\widehat{\beta}, \quad t = 1, \ldots, T. \qquad (8)$$

- Prediction observation: with subindex $p =$ (usually $p \notin [1, T]$):

$$Y_p = X_p'\beta + u_p. \qquad (9)$$

- Random disturbance u_p:

$$\mathrm{E}(u_p) = 0, \quad \mathrm{E}(u_p^2) = \sigma^2, \quad \mathrm{E}(u_p u_s) = 0 \quad \forall\, s \neq p.$$

- Known value of vector X_p' .

Point Prediction

- Substituting in SRF (8):

$$\boxed{\widehat{Y}_p = X_p' \widehat{\beta}.} \qquad (10)$$

i.e., numeric value as approximation to unknown value.

Prediction Error

- The error made (when taking $\widehat{Y}_p$ instead of the true Y_p) is

$$e_p = Y_p - \widehat{Y}_p,$$

- which can be expressed as:
- as a function of the two error sources introduced in the prediction.
- Under normality:

$$(\widehat{\beta} - \beta) \sim \mathcal{N}(0, \sigma^2 (X'X)^{-1}), \quad \text{and} \quad u_p \sim \mathcal{N}(0, \sigma^2),$$

- so that $\boxed{e_p \sim \mathcal{N}(0, \sigma_e^2)}$,
- where the prediction error variance is:

$$\sigma_e^2 = X_p' \underbrace{\text{Var}(\widehat{\beta})}_{\sigma^2 (X'X)^{-1}} X_p + \underbrace{\text{Var}(u_p)}_{\sigma^2} + \underbrace{\text{Cov}(\widehat{\beta}, u_p)}_{0}$$
$$= \sigma^2 (1 + X_p'(X'X)^{-1} X_p).$$

Interval Prediction

- Standardised prediction error:

$$\frac{e_p - 0}{\sigma_e} = \frac{e_p}{\sigma \sqrt{1 + X_p'(X'X)^{-1} X_p}} \sim \mathcal{N}(0,1),$$

- Recall how changing $\sigma \to \widehat{\sigma} \quad \Rightarrow \quad \mathcal{N}(0,1) \to t$!!, then

$$\frac{e_p}{\widehat{\sigma}_e} = \frac{e_p}{\widehat{\sigma} \sqrt{1 + X_p'(X'X)^{-1} X_p}} \sim t(T\text{–}K\text{–}1).$$

- Therefore:

$$Pr(-t_{\alpha/2} \leq \frac{e_p}{\widehat{\sigma}_e} \leq t_{\alpha/2}) = 1 - \alpha,$$

- and solving for Y_p :

$$Pr(\widehat{Y}_p - \widehat{\sigma}_e t_{\alpha/2} \leq Y_p \leq \widehat{Y}_p + \widehat{\sigma}_e t_{\alpha/2}) = 1 - \alpha.$$

- Then, the $(1-\alpha)$ confidence interval for the unknown Y_p is:

$$CI(Y_p)_{(1-\alpha)} = \left[\widehat{Y}_p \pm \widehat{\sigma}_e t_{\alpha/2} \right],$$

which measures the precision of the point prediction.

Prediction: Example

- In the previous example (linearised Cobb-Douglas fn:)

$$\widehat{\log Y}_t = \widehat{\alpha} + \widehat{\beta}_L \log L_t + \widehat{\beta}_K \log K_t, \quad T = 53;$$
$$\widehat{\log Y}_t = 2.10 + 0.67 \log L_t + 0.32 \log K_t, \quad \widehat{\sigma}^2 = 4$$

- "What value would Y_p take if $\log L_p = 2.5; \log K_p = 2.0$?":
- $X_p' = \begin{bmatrix} 1 & 2.5 & 2.0 \end{bmatrix}$
-

$$\widehat{\log Y_p} = X_p' \widehat{\beta} = \begin{bmatrix} 1 & 2.5 & 2.0 \end{bmatrix} \begin{bmatrix} 2.10 \\ 0.67 \\ 0.32 \end{bmatrix}$$
$$= 2.10 + 0.67 \cdot 2.5 + 0.32 \cdot 2.0 = 4.42$$

Prediction: Example

- Construct a 95% CI for the true Y_p:

$$\widehat{\sigma}_e^2 = \sigma^2(1 + X_p'(X'X)^{-1}X_p)$$

$$= 4\left(1 + \begin{bmatrix}1 & 2.5 & 2.0\end{bmatrix}\begin{pmatrix}2 & 0 & 0\\ 0 & 4 & -1\\ 0 & -1 & 7\end{pmatrix}\begin{bmatrix}1\\ 2.5\\ 2.0\end{bmatrix}\right)$$

$$= 4\left(1 + \begin{bmatrix}2 & 8 & 11.5\end{bmatrix}\begin{bmatrix}1\\ 2.5\\ 2.0\end{bmatrix}\right)$$

$$= 4\,(1 + 45) = 4 \cdot 46 = 184$$

-

$$CI(\log Y_p)_{0.95} = \left[\widehat{\log Y_p} \pm \widehat{\sigma}_e\, t_{0.025}(50)\right]$$

$$= \left[4.42 \pm \sqrt{184} \cdot 2.01\right]$$

$$= [4.42 \pm 27.25]$$

$$= [-22.84 \quad ; \quad 31.68]$$