INTRODUCTORY ECONOMETRICS

Lesson 3a

Dr Javier Fernández

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 $R = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow$

  $R \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \beta^\star$ (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

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

OLS residuals under Normality

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

  Then,

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

- 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"

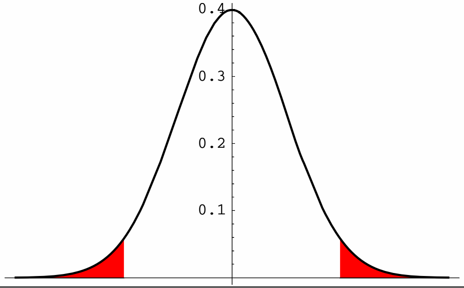
Hypothesis and Tests (rev2)

	elements	steps
a)	hypothesis to test (about estimator)	$H_0 : \dots$ vs. $H_a : \dots$ (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:

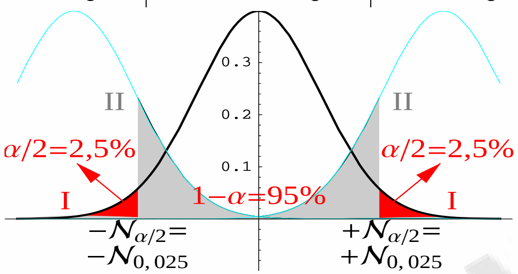
a)	$H_0 : \beta = 2.5$ vs. $H_a : \beta \neq 2.5$ (Var(β)=4)
b)	$\widehat{\beta} \sim \mathcal{N}(\beta, 4) \rightsquigarrow z = \frac{\widehat{\beta} - \beta}{2} \sim \mathcal{N}(0,1)$
c)	$z = \frac{\widehat{\beta} - 2.5}{2} \in$



Hypothesis and Tests: Critical region

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

critical region	confidence region	critical region

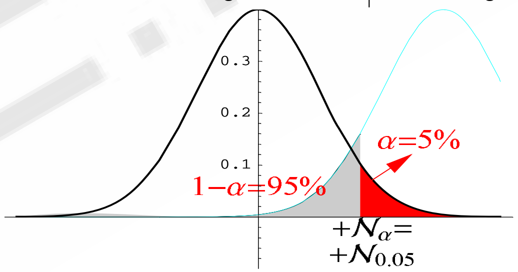


critical region	confidence region	critical region

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):

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

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

2. Def of t (Student): $\left.\begin{array}{c} Z \sim \mathcal{N}(0,1), \quad W \sim \chi^2(m) \\ Z, W \text{ independent} \end{array}\right\} \dfrac{Z}{\sqrt{W/m}} \sim t(m)$

3. Def of $\mathcal{F}$ (Snedecor): $\left.\begin{array}{c} V \sim \chi^2(n) \quad W \sim \chi^2(m) \\ V, W \text{ independent} \end{array}\right\} \dfrac{V/n}{W/m} \sim \mathcal{F}_m^n$

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

Hypothesis and Tests: Useful result

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

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

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 $\widehat{\sigma}$:

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

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

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 :

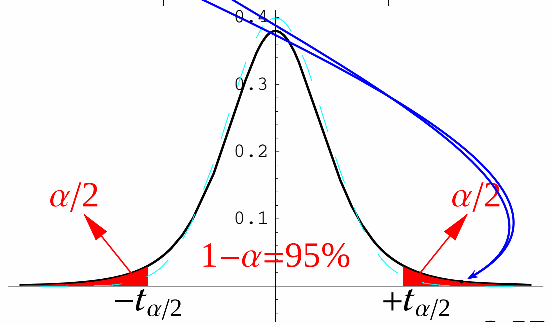
    $\Rightarrow$ β_i is (statistically or significantly) different from zero

    $\Rightarrow$ 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

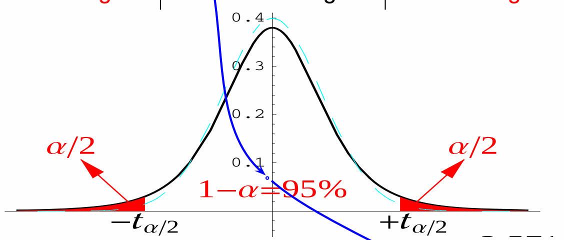


critical region	confidence region	critical region

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 $d_0, d_1, \dots, +d_K$) and
  $r = c$ (any single value):
- Let $H_0 : v = d'\beta = d_0\beta_0 + d_1\beta_1 + \cdots + d_K\beta_K = c$
  that is,
  an informative test about the value c that takes a single linear combination v 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 $c = 1$:

-

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

$$= c \qquad\qquad = 1$$

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

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

$$\widehat{v} \sim \mathcal{N}(v, \mathrm{Var}(\widehat{v}))$$

where $\mathrm{Var}(\widehat{v}) = \sigma^2 \sum_{i,j=0}^{K} d_i d_j a_{ij}$

- As before, standardise $\widehat{v}$

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

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

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

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

Testing a Single Linear Combination: rule

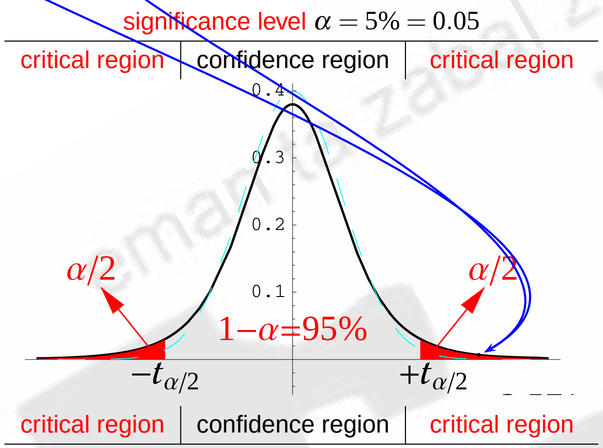
- $\boxed{\frac{\widehat{v} - v}{S_{\widehat{v}}} \sim t(T{-}K{-}1)}$
- Which Test? $\{ H_0 : v (= d'\beta) = c$ (informative test)
- **Remember:** Hypothesis $\rightsquigarrow$ statistic $\rightsquigarrow$ rule...
- Test for a linear combination:
  - Hypothesis: $H_0 : v = c$ vs. $H_a : v \neq c$
 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

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

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



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 (cont)

- 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 write Sum of Squares:
  - $$\frac{\widehat{\beta}^{\star\prime} x' x \widehat{\beta}^\star}{\sigma^2} \sim \chi^2(K) \quad \text{under } H_0 :$$
- 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

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

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

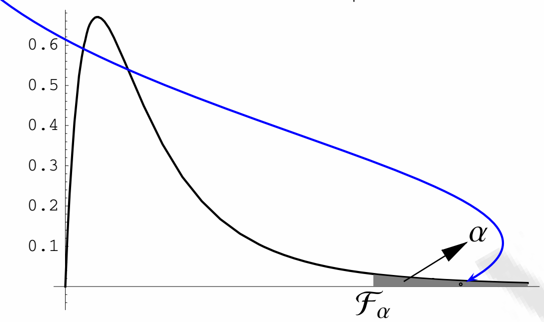
  - Rule: $F > \mathcal{F}_\alpha(K, T\text{–}K\text{–}1) \quad \Rightarrow \quad$ reject H_0 :
    - $\Rightarrow$ all coefs are jointly significant (different from zero)
    - $\Rightarrow$ whole regression is (statistically) relevant.

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.