Econometrics

Jerzy Mycielski

2013

Topics

- Statistical inference
- Most important distributions of random variables
- Assumptions of CNLRM
- Distributions of $\mathbf{b}$ and $\mathbf{e}$
- t statistics
- Simple and joint hypothesis, confidence intervals
- Joint hypothesis and F statistics
- Constrained minimization and second form of F statistics

Normal and related distributions

- Linear function (combination) of random variables with multivariate normal distribution has multivariate normal distribution
- $\mathbf{z} \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$
- $\mathbf{a}, \mathbf{B}$ nonrandom

$$a + \mathbf{Bz} \sim N\left(\mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B\Sigma B}'\right)$$

- If $\mathbf{z} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$ and $\boldsymbol{\Sigma}$ is nonsingular matrix $(r \times r)$, then

$$\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z} \sim \chi^2_r$$

- If $\mathbf{z} \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$ and matrix $\mathbf{A}$ is nonrandom idempotent matrix and $\mathrm{Rank}\left(\mathbf{A}\right) = r$, then

$$\frac{\mathbf{z}'\mathbf{Az}}{\sigma^2} \sim \chi^2_r$$

- if random variable $Z \sim N(0,1)$ and random variable $W \sim \chi^2_r$ and Z and W are independent, then

$$\frac{Z}{\sqrt{\frac{W}{r}}} \sim t_r$$

- If random variable $Z \sim \chi^2_k$ and random variable $W \sim \chi^2_r$ and Z and W are independent, then

$$\frac{\left(\frac{Z}{k}\right)}{\left(\frac{W}{r}\right)} \sim F\left(k, r\right)$$

Hypothesis testing and confidence intervals

- α significance level - probability of type I error (rejection of true H_0)
- k_α critical value for test statistics at significance level α

$$F\left(k_\alpha\right) = 1 - \alpha$$

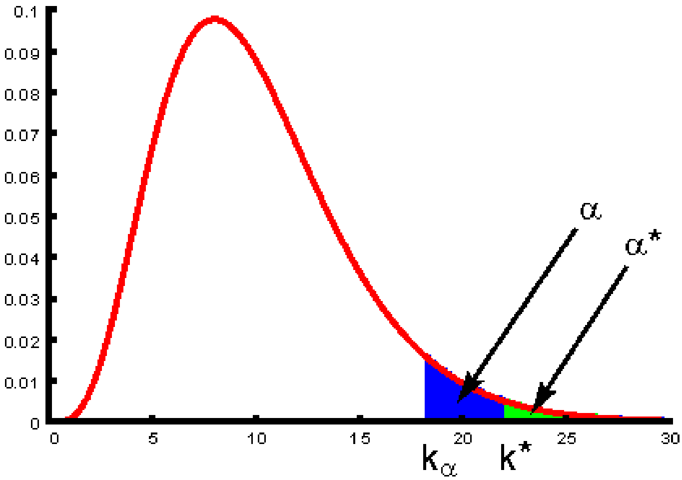
- k^* value of the test statistics in the sample
- α^* significance level calculated on the basis of the sample

$$\alpha^* = 1 - F\left(k^*\right)$$

- α^* is called p (p-*value*)

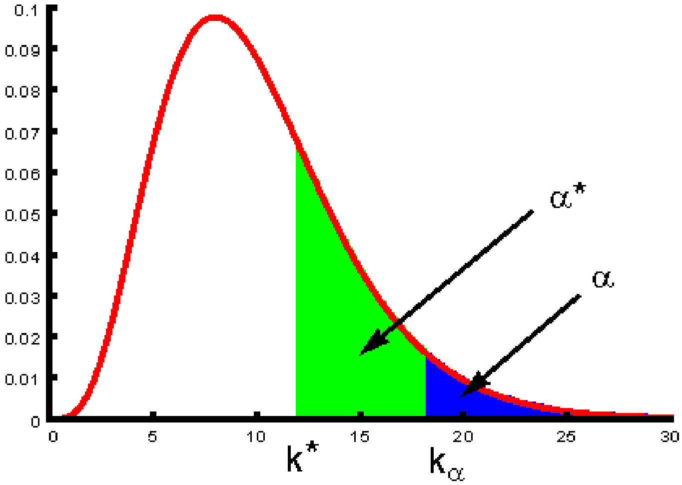
Value of k^*	Value of α^*	Decision
$k^* > k_\alpha$	$\alpha^* < \alpha$	Rejection of H_0
$k^* < k_\alpha$	$\alpha^* > \alpha$	Cannot reject H_0

Rejection of null hypothesis



$k^* > k_\alpha$, $\alpha^* < \alpha \Longrightarrow H_0$ rejected

Non rejection of null hypothesis



$k^* < k_\alpha$, $\alpha^* > \alpha \Longrightarrow H_0$ not rejected

Hypothesis testing and confidence intervals

Simple hypothesis

- Regression for Cobba-Douglas model

q	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
l	1.236623	.3801312	3.25	0.003	.4649154	2.00833
r	.0080266	.017226	0.47	0.644	-.026944	.0429972
t	.0106685	.0014495	7.36	0.000	.0077258	.0136112
_cons	-5.31012	3.678503	-1.44	0.158	-12.77788	2.157639

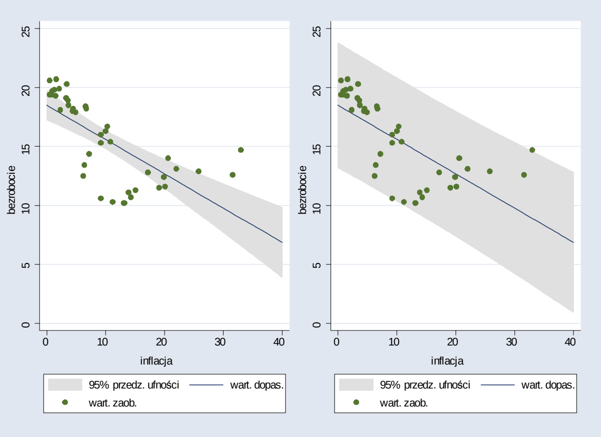
- Hypothesis of constant returns to scale $\beta_l = 1$

$$t = \frac{1.236623 - 1}{.3801312} = .622$$

Confidence intervals for forecast and for forecast error

- Confidence interval for forecast error based on Philips curve

$$17.323 - 2.523 \times 2.0226 = 12.220$$
$$17.323 + 2.523 \times 2.0226 = 22.426$$



Joint hypothesis testing

Example (Invalid formulation of joint hypothesis)

For $\beta = [\beta_1, \beta_2]'$, $\mathbf{H}\beta = \mathbf{h}$

$$\begin{cases} \beta_1 + \beta_2 = 1 \\ \beta_1 + \beta_2 = 2 \end{cases}$$

in this case $\mathbf{H} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $\mathbf{h} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

Formulating the joint hypothesis

- Model for the log of wages

$$\ln\left(\text{wage}_i\right) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \text{sex}_i + \boldsymbol{\beta}_2 \text{age}_i + \sum_{s=2}^{7} \gamma_s D_{s,i} + \boldsymbol{\varepsilon}_i$$

- **Hypothesis**: join insignificance of all variables apart form constant

$$\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = \gamma_6 = 0$$

- Test result: $R^2 = 0.2975$, $N = 6509$, $K = 9$:

$$F = \frac{6500}{8} \frac{0.2975}{1 - 0.2975} = 344.56$$

- p-value:

$$\boldsymbol{\alpha}^* = 1 - F^{-1}_{8,6500}\left(344.56\right) \approx 0.0000 < \boldsymbol{\alpha} = 0.05$$

Formulating the joint hypothesis

- **Hypothesis:** education is not significant

$$\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = \gamma_6 = 0$$

- Model with constraints

$$\ln(\text{wage}_i) = \beta_0 + \beta_1 \text{sex}_i + \beta_2 \text{age}_i + \varepsilon_i$$

- Test result: $S_R = 1771.145$. $S = 1342.156$.

$$F = \frac{(1771.145 - 1342.156)/6}{1342.156/6500} = 346.26$$

p-value:

$$\alpha^* = 1 - F^{-1}_{6,6500}(346.26) \approx 0.0000 < \alpha = 0.05$$

Formulating the joint hypothesis

- **Hypothesis**: only secondary and higher education is significant

$$\begin{cases} \gamma_6 = 0 \\ \gamma_5 = 0 \\ \gamma_4 = \gamma_3 \\ \gamma_3 = \gamma_2 \end{cases}$$

- Model with constraints

$$\begin{aligned} \ln(\text{wage}_i) &= \beta_0 + \beta_1 \text{sex}_i + \beta_2 \text{age}_i + \gamma_4 D_{4,i} \\ &\quad + \gamma_4 D_{5,i} + \gamma_4 D_{6,i} + \gamma_7 D_{7,i} + \varepsilon_i \end{aligned}$$

grouping the variables we obtain:

$$\begin{aligned} \ln(\text{wage}_i) &= \beta_0 + \beta_1 \text{sex}_i + \beta_2 \text{age}_i + \\ &\quad \gamma_4 (D_{4,i} + D_{5,i} + D_{6,i}) + \gamma_7 D_{7,i} + \varepsilon_i \end{aligned}$$

Formulating the joint hypothesis

- Define

$$\begin{aligned} \text{secondary}_i &= D_{4,i} + D_{5,i} + D_{6,i} \\ \text{higher}_i &= D_{7,i} \end{aligned}$$

resulting model

$$\ln(\text{wage}_i) = \beta_0 + \beta_1 \text{sex}_i + \beta_2 \text{age}_i + \gamma_4 \text{secondary}_i + \gamma_7 \text{higher}_i + \varepsilon_i$$

- $S_R = 1346.927$

$$F = \frac{(1346.927 - 1342.156)/4}{1342.156/6500} = 5.78$$

- p-value:

$$\alpha^* = 1 - F^{-1}_{4,6500}(5.78) \approx 0.0001 < \alpha = 0.05$$