Econometrics

Practical Session 10

Inference on Linear Combinations of Parameters

Ricardo Gouveia-Mendes
rgouveiamendes@ucp.pt

Spring 2025-26

Católica-Lisbon School of Business and Economics

Theoretical Wrap-up

Can We Test Restrictions With >1 Parameters?

- We already know how to test **restrictions involving only one parameter**: $H_0 : \beta_j = c$
- But what if we want to test **restrictions involving multiple parameters**? $H_0 : a\beta_j + b\beta_k = c$
- Notice that **defining** $\theta = a\beta_j + b\beta_k$ allows us to rewrite the restriction as $H_0 : \theta = c$

$$t = \frac{\hat{\theta} - c}{s_{\hat{\theta}}} \dot{\sim} \mathcal{N}(0, 1) \qquad \mathrm{CI}_{\theta}^{1-\alpha} = \left\{\hat{\theta} \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot s_{\hat{\theta}}\right\}$$

- **Problem: $s_{\hat{\theta}}$ is not directly available**

Can We Test Restrictions With >1 Parameters?

- If $\text{cov}\left(\hat{\beta}_j, \hat{\beta}_k\right) = 0 \longrightarrow s_{\hat{\theta}} = \sqrt{a^2 s^2_{\hat{\beta}_j} + b^2 s^2_{\hat{\beta}_k}}$
- Alternatively, notice that:

$$\theta = a\beta_j + b\beta_k \quad \Leftrightarrow \quad \beta_j = \frac{\theta}{a} - \frac{b}{a}\beta_k$$

$$Y = \beta_0 + \ldots + \left(\frac{\theta}{a} - \frac{b}{a}\beta_k\right) X_j + \beta_k X_k + \ldots + u \Leftrightarrow$$

$$Y = \beta_0 + \ldots + \theta \frac{X_j}{a} + \beta_k \left(X_k - X_j \cdot \frac{b}{a}\right) + \ldots + u$$

- Then, use R

What About Multiple Linear Restrictions?

- We can test **several linear restrictions at the same time**
- Usually, this means testing: $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$. Example:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$
$$Y = \beta_0 + \beta_1 X_1 + v$$

$$H_0 : \beta_2 = \beta_3 = 0 \qquad H_A : \neg H_0$$

What About Multiple Linear Restrictions?

- The statistic is:

$$F = \frac{\text{SSR}_* - \text{SSR}}{\text{SSR}} \frac{N-p}{q} = \frac{R^2 - R^2_*}{1 - R^2} \frac{N-p}{q} \sim F_{q,N-p}$$

  - SSR_* and R^2_* are from the restricted model
  - SSR and R^2 are from the unrestricted model
  - q is the number of restrictions
  - p is the number of parameters in the unrestricted model

Exercises

Stock & Watson – Exercise 7.5

Consider the following regression results, using 2012 data converted into 2007 units using the Consumer Price Index:

$$\widehat{\log(AHE)} = \underset{(0.049)}{12.471} + \underset{(0.021)}{0.373}\,\text{HighSchool} + \underset{(0.020)}{0.457}\,\text{Male} + \underset{(0.001)}{0.011}\,\text{Age}$$

$$\text{SER} = 1.023, \quad \bar{R}^2 = 0.0761$$

Consider also a new estimation for the same regression, but using 1993 data also converted to 2007 units:

$$\widehat{\log(AHE)} = \underset{(0.20)}{9.32} + \underset{(0.019)}{0.301}\,\text{HighSchool} + \underset{(0.047)}{0.562}\,\text{Male} + \underset{(0.002)}{0.011}\,\text{Age}$$

$$\text{SER} = 1.25, \quad \bar{R}^2 = 0.08$$

Stock & Watson – Exercise 7.5

Was there a statistically significant change in the coefficient on *HighSchool*?

- We want to test: $H_0 : \theta \equiv \beta_{1,2012} - \beta_{1,1993} = 0$ against $H_A : \theta \neq 0$
- To compute a t-statistic for θ we need to estimate $\text{var}\left(\hat{\theta}\right) \equiv \text{var}\left(\hat{\beta}_{1,2012} - \hat{\beta}_{1,1993}\right)$
- Since the two estimates come from different independent samples:

$$\text{var}\left(\hat{\theta}\right) = \text{var}\left(\hat{\beta}_{1,2012}\right) + \text{var}\left(\hat{\beta}_{1,1993}\right) \Rightarrow s_{\hat{\theta}} = \sqrt{\text{var}\left(\hat{\beta}_{1,2012}\right) + \text{var}\left(\hat{\beta}_{1,1993}\right)} = 0.02832$$

$$t = \frac{\hat{\theta} - 0}{s_{\hat{\theta}}} = 2.542373 > 1.96 = \Phi^{-1}(97.5\%)$$

- We reject H_0 at the 5% significance level: the effect of high school education on average hourly earnings has increased significantly

Stock & Watson – Exercise 7.7

Data were collected from a random sample of 200 home sales from a community in 2013. Let *Price* denote the selling price (in $1000s), *BDR* denote the number of bedrooms, *Bath* denote the number of bathrooms, *Hsize* denote the size of the house (in square feet), *Lsize* denote the lot size (in square feet), *Age* denote the age of the house (in years), and *Poor* denote a binary variable that is equal to 1 if the condition of the house is reported as “poor.”

Stock & Watson – Exercise 7.7

An estimated regression yields:

$$\widehat{Price} = \underset{(22.1)}{109.7} + \underset{(1.23)}{0.567}\,BDR + \underset{(9.76)}{26.9}\,Bath + \underset{(0.021)}{0.239}\,Hsize + \underset{(0.00072)}{0.005}\,Lsize$$

$$+\underset{(0.23)}{0.1}\,Age - \underset{(12.23)}{56.9}\,Poor, \qquad \bar{R}^2 = 0.85, \quad SER = 45.8$$

a) Is the coefficient on *BDR* statistically significantly different from zero?

$$H_0 : \beta_1 = 0 \quad H_1 : \beta_1 \neq 0$$

$$t = \frac{0.567 - 0}{1.23} = 0.46098 < 1.96 = \Phi^{-1}(97.5\%)$$

We do not reject H_0, so the coefficient on *BDR* is not statistically significantly different from zero.

Stock & Watson – Exercise 7.7

b) Typically, four-bedroom houses sell for more than three-bedroom houses. Is this consistent with your answer to a) and with the regression more generally?

> - The coefficient on *BDR* is positive, pointing to a higher price the higher the number of bedrooms
> - However it is not statistically significant, so we cannot conclude that four-bedroom houses sell for more than three-bedroom houses

Stock & Watson – Exercise 7.7

c) A homeowner purchases 2500 square feet from an adjacent lot. Construct a 95% confident interval for the change in the value of her house.

- The estimator can be written as: $\theta = 2500\,\beta_4$
- To build a confidence interval we need: $\mathrm{CI}_{\theta}^{95\%} = \left\{\hat{\theta} \pm s_{\hat{\theta}} \cdot \Phi^{-1}(97.5\%)\right\}$

$$\hat{\theta} = 2500 \cdot 0.005 = 12.5 \qquad s_{\hat{\theta}} = \sqrt{2500^2 \cdot s_{\hat{\beta}_4}^2} = 2500 \cdot s_{\hat{\beta}_4} = 1.8$$

$$\mathrm{CI}_{\theta}^{95\%} = [8.972, 16.028]$$

Stock & Watson – Exercise 7.7

d) Lot size is measured in square feet. Do you think that another scale might be more appropriate? Why or why not?

> - The scale of the variables should make easy to read the regression results
> - If the lot size were measured in thousands of square feet, the estimate coefficient would be 5 instead of 0.005

Stock & Watson – Exercise 7.7

e) The F-statistic for omitting *BDR* and *Age* from the regression is $F = 2.38$. Are the coefficients on *BDR* and *Age* statistically different from zero at the 10% level?

> - $F^{-1}_{2,193}(90\%) = 2.33 < 2.38$
> - So we reject the null hypothesis that the coefficients on *BDR* and *Age* are zero at the 10% level
> - The coefficients are jointly significant