Testing Multiple Linear Restrictions

Econometrics

Péter Elek

Analysis of Variance (ANOVA) Table

- $H_0 : \beta_j = 0 \quad (j = 1, \ldots, k)$ i.e. only the constant is present in the regression

Source of Variation	**Expected Value under H_0**	**Mean Sum of Square**	**F**
Explained (SSE)	$k\sigma^2$	$SSE/k = MS_1$	$F = MS_1/MS_2$
Residual (SSR)	$(n-k-1)\sigma^2$	$SSR/(n-k-1) = MS_2$	
Total (SST)	$(n-1)\sigma^2$		

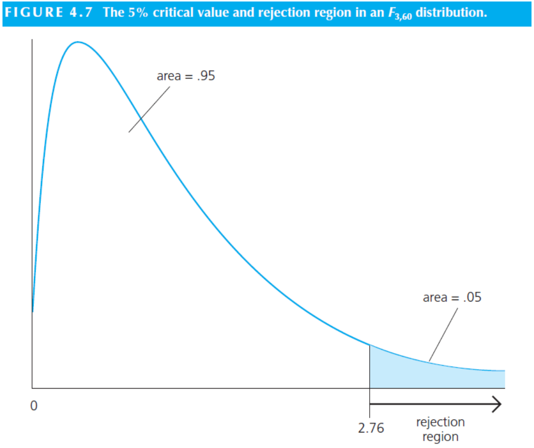
F-test for the Overall Significance of a Regression $H_0 : \beta_j = 0 \ (j = 1, \ldots, k)$

- $$F = \frac{SSE/k}{SSR/(n-k-1)} = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$
- Suppose assumptions (1)-(4) and homoskedasticity of the error term holds.
- Then, in large samples, under H_0,

$$F \sim F_{k,n-k-1}$$

- (F distribution with degrees of freedom k and $n-k-1$)
- Reject H_0 if $F >$ critical value from the $F_{k,n-k-1}$ distribution
- (One-tailed test because $F > 0$ and the positive deviation from zero is tested.)

Example: F Distribution (always a nonnegative-valued distribution)



Source: Wooldridge Fig. 4.7.

F-test for General Linear Restrictions

- Testing r restrictions in a regression with k explanatory variables
- Nested models: unrestricted model includes the restricted model as a special case
- Example: exclusion restriction
  - Unrestricted: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$
  - $H_0 : \beta_2 = 0$ and $\beta_3 = 0$
  - Restricted: $y = \beta_0 + \beta_1 x_1 + v$
- But the restriction can be more general, e.g.
  - Unrestricted: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$
  - $H_0 : \beta_2 = 1$ and $\beta_3 = 1$
  - Restricted: $y = \beta_0 + \beta_1 x_1 + x_2 + x_3 + v$

F-test for General Linear Restrictions (cont.)

- $$F = \frac{(SSR_r - SSR_u)/r}{SSR_u/(n-k-1)} = \frac{(R_u^2 - R_r^2)/r}{(1-R_u^2)/(n-k-1)}$$
- where:
  - SSR_r and R_r^2: sum of squared residuals and R-squared of restricted model
  - SSR_u and R_u^2: sum of squared residuals and R-squared of unrestricted model
  - r: number of restrictions
  - k : number of parameters in the unrestricted model (excluding the constant)
  - n : number of observations
- (Note: can only be used if the sample size and the dependent variable are the same in the U and R model, e.g. there are no missing values.)
- Under H_0 :
  $$F \sim F_{r,n-k-1}$$
- We reject H_0 if $F >$ critical value of the $F_{r,n-k-1}$ distribution

Example: Black-White Wage Differences

TABLE 1

LOG WAGE REGRESSIONS BY SEX

	MEN (N = 1,593)			WOMEN (N = 1,446)		
	(1)	(2)	(3)	(4)	(5)	(6)
Black	−.244	−.196	−.072	−.185	−.155	.035
	(.026)	(.025)	(.027)	(.029)	(.027)	(.031)
Hispanic	−.113	−.045	.005	−.028	.057	.145
	(.030)	(.029)	(.030)	(.033)	(.031)	(.032)
Age	.048	.046	.040	.010	.009	.023
	(.014)	(.013)	(.013)	(.015)	(.014)	(.015)
AFQT	...	...	.172	...	...	.228
			(.012)			(.015)
AFQT^2	...	...	−.013	...	...	.013
			(.011)			(.013)
High grade by 1991	...	.061	...	...	.088	...
		(.005)			(.005)	
R^2	.059	.155	.168	.029	.191	.165

Source: Neal and Johnson (1996): The Role of Premarket Factors in Black-White Wage Differences. Journal of Political Economy 104, 869-895.
AFQT: Armed Forces Qualification Test (taken at teenage years)

Example: F-tests (Sample for Men)

- $H_0 : \beta_{black} = \beta_{hisp} = \beta_{age} = 0$ in column (1)
    - $F = \frac{0.059/3}{(1-0.059)/(1593-3-1)} = 33.2$
    - 99.9% quantile of $F_{3,1589}$ is $\text{qf}(0.999, 3, 1589) = 5.5$
    - so H_0 is rejected at all reasonable significance levels (not surprisingly)
- $H_0 : \beta_{AFQT} = \beta_{AFQT2} = 0$ (test between column (3) and (1))
    - $F = \frac{(0.168-0.059)/2}{(1-0.168)/(1593-3-1)} = 104.1$
    - 99.9% quantile of $F_{2,1589}$ is $\text{qf}(0.999, 2, 1589) = 6.9$
    - so H_0 is rejected at all reasonable significance levels (not surprisingly)

Testing a Single Linear Combination of Parameters

- Example: Cobb-Douglas production function

$$\log Y = \beta_0 + \beta_1 \log L + \beta_2 \log K + u$$

- A relevant null hypothesis:

$$H_0 : \beta_1 + \beta_2 = 1$$

Testing a Single Linear Combination of Parameters (cont.)

- Method 1: reparametrization: $\theta = \beta_1 + \beta_2$, and test $H_0 : \theta = 1$ in

$$\log Y = \beta_0 + \theta \log L + \beta_2(\log K - \log L) + u$$

- Method 2: t-test for the linear combination by calculating the variance

$$t = \frac{\hat{\beta}_1 + \hat{\beta}_2 - 1}{SE(\hat{\beta}_1 + \hat{\beta}_2)}$$

$$\text{Var}(\hat{\beta}_1 + \hat{\beta}_2) = \text{Var}(\hat{\beta}_1) + \text{Var}(\hat{\beta}_2) + 2\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$$

- Method 3: F-test by imposing restriction $\beta_1 = 1 - \beta_2$ and comparing the restricted and unrestricted model
- In-built commands conduct such tests in statistical softwares.

Large Sample Heteroskedasticity-Robust Tests

- In large samples, F-test is valid under assumptions (1)-(4) and homoskedasticity
- Other large sample tests exist: Wald-test and LM-test (Lagrange multiplier test)
- Heteroskedasticity-robust versions exist and are available in software packages

Remark: Testing for Heteroskedasticity

- In large sample parameter estimation, homoskedasticity is usually not important (robust inference exists, and usually does not differ much from conventional one).

- If needed: $H_0 : Var(u \mid \mathbf{X}) = \sigma^2$
- Idea: under H_0, u^2 does not depend on any combination of x-s
- So: (1) Obtain $\hat{u}_i$ from the main regression
- (2) Auxiliary regression: regress $\hat{u}_i^2$ on
  - $x_{1i}, \ldots, x_{ki}$ (Breusch-Pagan test)
  - or on the explanatory variables, their squares and their cross-products (White-test)
  - or on the $\hat{y}_i$ predicted value and its square $\hat{y}_i^2$ (special form of White-test)
- (3) Then test the null hypothesis that none of the variables is nonzero in the auxiliary regression.

Material

- Wooldridge ch 4, 5.1-5.2 (excluding LM test), 8.2-8.3 (excluding robust LM test)