Elements of Econometrics.

Lecture 5. Multicollinearity. Linear Restrictions.

FCS, 2022-2023

MULTICOLLINEARITY

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u \qquad\qquad X_3 = \lambda + \mu X_2$$

$$\hat{\beta}_2 = \frac{\widehat{\mathrm{Cov}}(X_2, Y)\widehat{\mathrm{Var}}(X_3) - \widehat{\mathrm{Cov}}(X_3, Y)\widehat{\mathrm{Cov}}(X_2, X_3)}{\widehat{\mathrm{Var}}(X_2)\widehat{\mathrm{Var}}(X_3) - \left[\widehat{\mathrm{Cov}}(X_2, X_3)\right]^2} =$$

$$= \frac{\widehat{\mathrm{Cov}}(X_2, Y)\widehat{\mathrm{Var}}(\lambda + \mu X_2) - \widehat{\mathrm{Cov}}(\lambda + \mu X_2, Y)\widehat{\mathrm{Cov}}(X_2, \lambda + \mu X_2)}{\widehat{\mathrm{Var}}(X_2)\widehat{\mathrm{Var}}(\lambda + \mu X_2) - \left[\widehat{\mathrm{Cov}}(X_2, \lambda + \mu X_2)\right]^2} =$$

$$= \frac{\mu^2\widehat{\mathrm{Cov}}(X_2, Y)\widehat{\mathrm{Var}}(X_2) - \mu^2\widehat{\mathrm{Cov}}(X_2, Y)\widehat{\mathrm{Cov}}(X_2, X_2)}{\mu^2\widehat{\mathrm{Var}}(X_2)\widehat{\mathrm{Var}}(X_2) - \mu^2\left[\widehat{\mathrm{Cov}}(X_2, X_2)\right]^2} = \frac{0}{0}$$

$$\mathrm{s.e.}(\hat{\beta}_2) = s_u \times \frac{1}{\sqrt{n}} \times \frac{1}{\sqrt{\Sigma {x_{2i}}^2 / n}} \times \frac{1}{\sqrt{1 - r^2_{X_2, X_3}}}$$

$$r^2_{X_2, X_3} = 1 \Rightarrow \sqrt{1 - r^2_{X_2, X_3}} = 0$$

$X_3 = \lambda + \mu X_2$ - Perfect multicollinearity. The effects of X_2 and X_3 can not be separated.

No OLS estimation can be done.

POSSIBLE MEASURES FOR ALLEVIATING MULTICOLLINEARITY

$$\sigma^2_{\hat{\beta}_2} = \frac{\sigma_u^2}{\sum(X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r^2_{X_2,X_3}} = \frac{\sigma_u^2}{n\widehat{\mathrm{Var}}(X_2)} \times \frac{1}{1 - r^2_{X_2,X_3}}$$

(1) Reduce σ_u^2 by including further relevant variables in the model.

(2) Increase the number of observations.

(3) Increase $\widehat{\mathrm{Var}}(X_2)$.

(4) Reduce r_{X_2,X_3}.

(5) Drop some of the correlated variables.

(6) Combine the correlated variables

(7) Impose a restriction

LINEAR RESTRICTION: EDUCATIONAL ATTAINMENT FUNCTION EXAMPLE

S – years of schooling;
SM – years of schooling of mother;
SF – years of schooling of father.

If SM and SF are strongly correlated, the coefficients may be insignificant due to multicollinearity.

Options for a restriction: 1) $\beta_3 = 0;$ **2)** $\beta_4 = 0;$ **3)** $\beta_3 = \beta_4$.

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 SM + \beta_4 SF + u$$

$$\beta_3 = \beta_4$$

$$\begin{aligned} S &= \beta_1 + \beta_2 ASVABC + \beta_3 (SM + SF) + u \\ &= \beta_1 + \beta_2 ASVABC + \beta_3 SP + u \end{aligned}$$

Here we define *SP* as the sum of *SM* and *SF* (total parental schooling as the indicator of family background). The problem caused by multicollinearity has been eliminated.

LINEAR RESTRICTION: UoL 2020 Exams Results

Sample of 60 ICEF Year 2 students was taken randomly. UL_AVE – mean grade of 2020 UoL Exams; RATING_1S – Rating for Semester 1.

Dependent Variable: UL_AVE
Method: Least Squares
Date: 09/30/20 Time: 16:03
Sample: 1 65
Included observations: 51

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	67.43972	2.676749	25.19464	0.0000
MIC_1S	0.047118	0.101921	0.462295	0.6461
STATS_1S	0.075190	0.086343	0.870821	0.3885
MATH_1S	0.211854	0.107375	1.973031	0.0547
BANK_1S	-0.192209	0.106760	-1.800384	0.0785
ACC_1S	0.129469	0.087425	1.480909	0.1456

R-squared	0.404407	Mean dependent var	79.97255
Adjusted R-squared	0.338231	S.D. dependent var	7.414582
S.E. of regression	6.031704	Akaike info criterion	6.542067
Sum squared resid	1637.166	Schwarz criterion	6.769341
Log likelihood	-160.8227	Hannan-Quinn criter.	6.628915
F-statistic	6.111002	Durbin-Watson stat	2.184320
Prob(F-statistic)	0.000212		

Dependent Variable: UL_AVE
Method: Least Squares
Date: 09/30/20 Time: 16:04
Sample: 1 65
Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	68.76379	2.621290	26.23281	0.0000
RATING_1S	0.250972	0.056674	4.428377	0.0000

R-squared	0.252679	Mean dependent var	79.64556
Adjusted R-squared	0.239794	S.D. dependent var	8.108032
S.E. of regression	7.069377	Akaike info criterion	6.782187
Sum squared resid	2898.614	Schwarz criterion	6.851998
Log likelihood	-201.4656	Hannan-Quinn criter.	6.809494
F-statistic	19.61053	Durbin-Watson stat	2.574148
Prob(F-statistic)	0.000043		

Left: ICEF Grades in Semester 1 are Individually Insignificant due to Multicollinearity.

Rating (average grade) in Semester 1 is Significant at 0.1% Significance Level.

F TEST FOR TESTING THE HYPOTHESIS OF EQUALITY OF ALL SLOPE COEFFICIENTS TO 0

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

$H_0: \beta_2 = \ldots = \beta_k = 0$: (*k-1*) linear restrictions

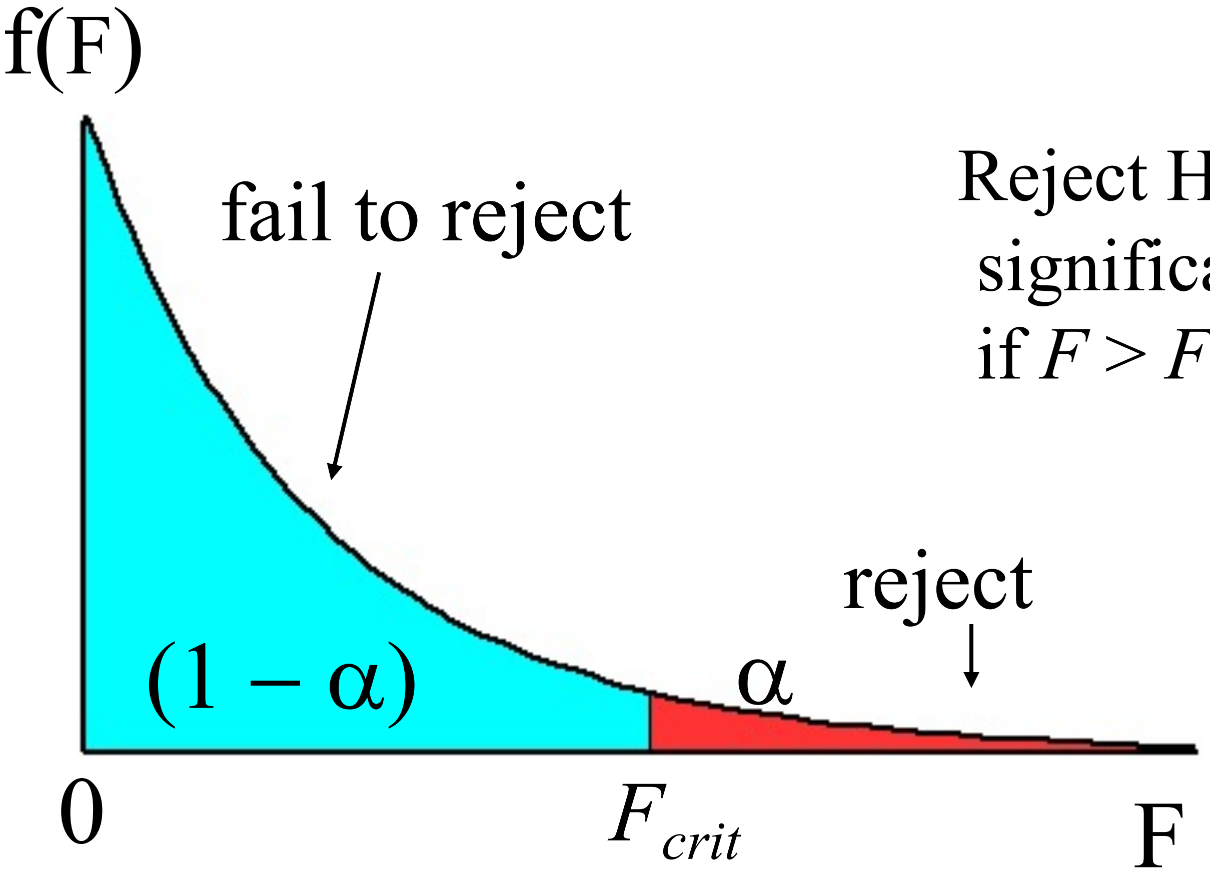
H_1: at least one of slope $\beta's \neq 0$

$$F(k-1, n-k) = \frac{(SSR_r - SSR_{ur})/(k-1)}{SSR_{ur}/(n-k)} = \frac{(SST - SSR)/(k-1)}{SSR/(n-k)} =$$

$$= \frac{SSE/(k-1)}{SSR/(n-k)} = \frac{\frac{SSE}{SST}/(k-1)}{\frac{SSR}{SST}/(n-k)} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

$$F\text{ (cost in d.f., d.f. unrestricted)} = \frac{\text{reduction in } SSR \;/\; \text{cost in d.f.}}{SSR \text{ unrestricted} \;/\; \text{degrees of freedom unrestricted}}$$

The *F* test and *F* statistic (continued)



Reject H_0 at α significance level if $F > F_{crit}$

F TESTS RELATING TO GROUPS OF EXPLANATORY VARIABLES

$$Y = \beta_1 + \beta_2 X_2 + u \qquad SSR_1$$

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u \qquad SSR_2$$

$$H_0: \beta_3 = \beta_4 = 0$$

$$H_1: \beta_3 \neq 0 \text{ or } \beta_4 \neq 0 \text{ or both } \beta_3 \text{ and } \beta_4 \neq 0$$

F **(cost in d.f., d.f. unrestricted) =** $\dfrac{\textbf{reduction in } \textbf{\textit{SSR}} \;/\; \textbf{cost in d.f.}}{\textbf{\textit{SSR}} \textbf{ unrestricted} \;/\; \textbf{degrees of freedom unrestricted}}$

$$F(2, n-4) = \frac{(SSR_1 - SSR_2)/2}{SSR_2/(n-4)} = \frac{(R_2^2 - R_1^2)/2}{(1 - R_2^2)/(n-4)}$$

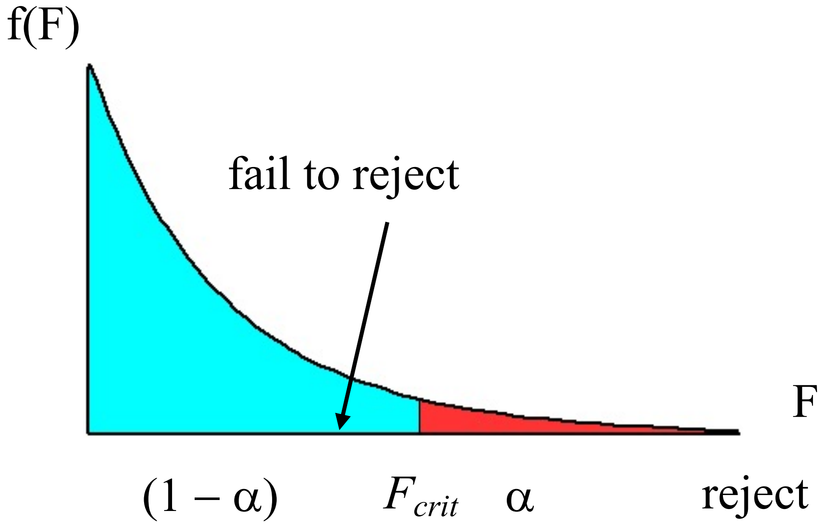
F TESTS FOR LINEAR RESTRICTIONS IN GENERAL, AND FOR GROUPS OF EXPLANATORY VARIABLES

$$F\,(\text{cost in d.f., d.f. unrestricted}) = \frac{\text{reduction in } SSR \;/\; \text{cost in d.f.}}{SSR \text{ unrestricted} \;/\; \text{degrees of freedom unrestricted}}$$

$$F(q, n-k) = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k)} = \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/(n-k)},$$

$(r\ -\ restricted, ur\ -\ unrestricted.$

$q - the\ number\ of\ linear\ restrictions\ on\ coefficients).$



EXAMPLE OF F-TESTS FOR LINEAR RESTRICTIONS: EARNINGS FUNCTION

Dependent Variable: EARN
Method: Least Squares
Included observations: 570

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-25.12985	5.004870	-5.021080	0.0000
S	1.791735	0.252261	7.102694	0.0000
ASVABC	0.128165	0.051480	2.489636	0.0131
MALE	4.083965	0.719913	5.672861	0.0000
PWE	0.420852	0.168994	2.490332	0.0130
ETHBLACK	-2.077643	1.334459	-1.556917	0.1201
ETHHISP	-2.707987	1.583041	-1.710623	0.0877

R-squared	0.242607	Mean dependent var	13.68988
S.D. dependent var	9.702960	S.E. of regression	8.489199
Sum squared resid	40573.44	F-statistic	30.05662
Durbin-Watson stat	1.879758	Prob(F-statistic)	0.000000

$$\widehat{EARNINGS} = -25.13 + 1.79S + 0.13ASVABC + 4.08MALE + 0.42PWE$$
$$-2.08ETHBLACK - 2.71ETHHISP$$

EXAMPLE OF F-TESTS FOR LINEAR RESTRICTIONS: EARNINGS FUNCTION

Unrestricted: $\widehat{EARNINGS} = \hat{\beta}_1 + \hat{\beta}_2 S + \hat{\beta}_3 ASVABC +$
$+\hat{\beta}_4 MALE + \hat{\beta}_5 PWE + \hat{\beta}_6 ETHBLACK + \hat{\beta}_7 ETHHISP$ $\qquad SSR_{ur} = 40573.44$

Restricted model H_0: $\beta_6 = \beta_7 = 0$

Dependent Variable: EARN Method: Least Squares Included observations: 570

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-26.32962	4.984233	-5.282582	0.0000
S	1.723689	0.249993	6.894949	0.0000
ASVABC	0.165788	0.048232	3.437296	0.0006
MALE	4.105127	0.720810	5.695156	0.0000
PWE	0.412816	0.168736	2.446514	0.0147

R-squared	0.236233	Mean dependent var	13.68988
S.D. dependent var	9.702960	S.E. of regression	8.509745
Sum squared resid	40914.91	F-statistic	43.68865
Durbin-Watson stat	1.881990	Prob(F-statistic)	0.000000

$$F(2{,}570-7) = \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/(570-7)} = \frac{(40914.91 - 40573.44)/2}{40573.44/563} = 4.74$$

$F_{\text{crit, 1\%}}(2;563) = 4.64$ $\qquad H_0\colon \beta_6 = \beta_7 = 0 \quad \textit{rejected}$

EXAMPLE OF F-TESTS FOR LINEAR RESTRICTIONS: EARNINGS FUNCTION

$$\widehat{EARNINGS} = \hat{\beta}_1 + \hat{\beta}_2 S + \hat{\beta}_3 ASVABC + \hat{\beta}_4 MALE + \hat{\beta}_5 PWE + \hat{\beta}_6 ETHBLACK + \hat{\beta}_7 ETHHISP - UR$$

$$\widehat{EARNINGS} = \hat{\beta}_1 + \hat{\beta}_2 S + \hat{\beta}_3 ASVABC + \hat{\beta}_4 MALE + \hat{\beta}_5 PWE + \hat{\beta}_6 ETHBLACK + \hat{\beta}_6 ETHHISP - Restricted$$

$$H_0: \beta_6 = \beta_7 \qquad ETHNW = ETHBLACK + ETHHISP$$

$$H_1: \beta_6 \neq \beta_7$$

Dependent Variable: EARN   Method: Least Squares   Included observations: 570

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-25.20476	4.995630	-5.045362	0.0000
S	1.797148	0.251515	7.145295	0.0000
ASVABC	0.127068	0.051329	2.475574	0.0136
MALE	4.073660	0.718649	5.668501	0.0000
PWE	0.425351	0.168296	2.527395	0.0118
ETHNW	-2.331664	1.082538	-2.153886	0.0317

R-squared	0.242464	Mean dependent var	13.68988
S.D. dependent var	9.702960	S.E. of regression	8.482471
Sum squared resid	40581.10	F-statistic	36.10389
Durbin-Watson stat	1.880030	Prob(F-statistic)	0.000000

EXAMPLE OF F-TESTS FOR LINEAR RESTRICTIONS: EARNINGS FUNCTION

$$\widehat{EARNINGS} = \hat{\beta}_1 + \hat{\beta}_2 S + \hat{\beta}_3 ASVABC + \hat{\beta}_4 MALE + \hat{\beta}_5 PWE + \hat{\beta}_6 ETHNW \quad - Restricted$$

$$\widehat{EARNINGS} = -25.2 + 1.80S + 0.13ASVABC + 4.07MALE + 0.43PWE - 2.33ETHNW$$

$$F(1{,}570-7) = \frac{(SSR_R - SSR_{UR})/1}{SSR_{UR}/(570-7)} = \frac{(40581.10 - 40573.44)/1}{40573.44/563} = 0.11$$

$$F_{\text{crit, }0.1\%}(1;563) = 10.95$$

$$H_0\colon \beta_6 = \beta_7 \quad not \quad rejected$$

t-TESTS FOR LINEAR RESTRICTIONS

$$\widehat{EARNINGS} = \hat{\beta}_1 + \hat{\beta}_2 S + \hat{\beta}_3 ASVABC + \hat{\beta}_4 MALE + \hat{\beta}_5 PWE + \hat{\beta}_6 ETHBLACK + \hat{\beta}_7 ETHHISP$$

$$H_0: \beta_6 = \beta_7$$
$$H_1: \beta_6 \neq \beta_7$$

$$H_0: (\beta_7 - \beta_6) = 0$$
$$H_1: (\beta_7 - \beta_6) \neq 0$$

$$\widehat{EARNINGS} = \hat{\beta}_1 + \hat{\beta}_2 S + \hat{\beta}_3 ASVABC + \hat{\beta}_4 MALE + \hat{\beta}_5 PWE + \hat{\beta}_6 (ETHBLACK + ETHHISP) + \\ + (\hat{\beta}_7 - \hat{\beta}_6) ETHHISP$$

or

$$\widehat{EARNINGS} = \hat{\beta}_1 + \hat{\beta}_2 S + \hat{\beta}_3 ASVABC + \hat{\beta}_4 MALE + \hat{\beta}_5 PWE + \hat{\beta}_6 ETHNW + (\hat{\beta}_7 - \hat{\beta}_6) ETHHISP$$

t-TESTS FOR LINEAR RESTRICTIONS: EXAMPLES

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \; ; \; i = 1, 2, \ldots, n \qquad (1)$$

1. $\beta_2 = 1$: $\quad Y_i = \beta_1 + X_{2i} + (\beta_2 - 1)X_{2i} + \beta_3 X_{3i} + u_i \;\rightarrow$

$$Y_i - X_{2i} = \beta_1 + (\beta_2 - 1)X_{2i} + \beta_3 X_{3i} + u_i \qquad (2)$$

Fit the regression and test H_0: $\beta_2 - 1 = 0$.

2. $\beta_2 + \beta_3 = 1$: $\; Y_i = \beta_1 + (\beta_2 + \beta_3 - 1)X_{2i} + (-\beta_3 + 1)X_{2i} + \beta_3 X_{3i} + u_i \rightarrow$

$$Y_i - X_{2i} = \beta_1 + (\beta_2 + \beta_3 - 1)X_{2i} + \beta_3(X_{3i} - X_{2i}) + u_i \qquad (3)$$

Fit the regression and test H_0: $\beta_2 + \beta_3 - 1 = 0$