

Multicollinearity

$$Y_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_k X_{kt} + U_t$$

$$E(U_t) = 0$$

$$\text{Var}(U_t) = \sigma^2 \text{ (homoscedasticity, no heteroscedasticity)}$$

$$\text{Cov}(U_t, U_s) = 0 \text{ (no autocorrelation, no serial correlation)}$$

and

there is No perfect linear relationship between the explanatory variables

No perfect multicollinearity

↳ for exp: $x_i = a + b x_i$ is ruled out

Consider

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + U_t$$

If there is a perfect relationship between X_{1t} & X_{2t} then

$$\hat{\beta} = (X'X)^{-1} X'Y$$

$(X'X) \Rightarrow$ singular

$$V(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

$$\hat{\sigma}_{\hat{\beta}_1}^2 = \frac{\hat{\sigma}^2}{\sum \hat{u}_{ti}^2} = \frac{\hat{\sigma}^2}{\sum x_{ti}^2 (1 - r_{12}^2)}$$

$$\text{If } r_{12}^2 = 1 \Rightarrow \hat{\sigma}_{\hat{\beta}_1}^2 : \text{Indeterminate}$$

example:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + U_t$$

let

$$X_{t1} = \delta_2 X_{t2}$$

$$\Rightarrow Y_t = \beta_0 + \beta_1 (\delta_2 X_{t2}) + \beta_2 X_{t2} + U_t$$

↳ cannot be estimated

re-arranging

$$Y_t = \beta_0 + \underbrace{(\beta_1 \delta_2 + \beta_2)}_{\alpha} X_{t2} + U_t$$

$$Y_t = \beta_0 + \alpha X_{t2} + U_t \text{ can be estimated}$$

BUT $\alpha = \beta_1 \delta_2 + \beta_2$

$\Rightarrow$ No unique value for the parameters

there may be imperfect relationship between the explanatory variables

This does not violate any of the basic assumptions:
the estimators are still BLUE

examples:

crop yield = $f(\text{rainfall, temperature})$

output = $f(\text{labour, capital})$

grade = $f(\text{attendance, study})$

The Consequences of Serious Multicollinearity

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

$$\hat{\beta}_i = \frac{\sum \hat{u}_{ti} y_t}{\sum \hat{u}_{ti}^2}$$

$$\hat{\sigma}_{\hat{\beta}_i}^2 = \frac{\hat{\sigma}^2}{\sum \hat{u}_{ti}^2}$$

for $k=2$:

$$\hat{\sigma}_{\hat{\beta}_i}^2 = \frac{\hat{\sigma}^2}{\sum x_{ti}^2 (1 - r_{12}^2)}$$

VIF

Let P_i^2 : the coef. of determination from the regression of i^{th} expl. variable on the rest of the variables

$$P_i^2 = 1 - \frac{\sum \hat{u}_{ti}^2}{\sum x_{ti}^2}$$

$$\Rightarrow \sum \hat{u}_{ti}^2 = \sum x_{ti}^2 (1 - P_i^2)$$

$$\hat{\sigma}_{\hat{\beta}_i}^2 = \frac{\hat{\sigma}^2}{\sum \hat{u}_{ti}^2} = \frac{\hat{\sigma}^2}{\sum x_{ti}^2} \cdot \boxed{\frac{1}{1 - P_i^2}}$$

VIF

Variance Inflation Factor (VIF)

$$VIF = \frac{1}{1 - R_i^2}$$

R_i^2	VIF	$\text{Var}(\hat{\beta}_i)$
0.0	1.0	A
0.5	2.0	2A
0.9	10.0	10A
0.95	50	50A
0.99	100	100A

$\Rightarrow t_{\hat{\beta}_i} = \frac{\hat{\beta}_i}{\sigma_{\hat{\beta}_i}}$

$\sigma_{\hat{\beta}_i}$

→ CONTAMINATED

→ smaller t-ratios

→ incorrectly do not reject H_0

higher VIF: Prob. of contamination
 power of test decreases

Multicollinearity: Serious?

i) $VIF_i \geq 10$ An indicator of serious MC

ii) t & F tests:

high $R^2 \rightarrow$ high F value: the variables are jointly insignificant

$\Rightarrow$ an indicator of serious MC

R_x :

R_x : the correlation matrix of expl. variables

$$k=2 \quad R_x = \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}$$

$$|R_x| = 1 - r_{12}^2$$

$$\text{If } r_{12}^2 = 1 \Rightarrow |R_x| = 0$$

$$r_{12}^2 = 0 \Rightarrow |R_x| = 1$$

If $|R_x| \rightarrow 0$: serious multicollinearity

Solutions to serious MC

i- delete the variable? Which contributes most?

But: Omitted variable: The coeffs of var. remain
expl. variables will be Biased

ii) Alternative elimination methods?

→ Principal components etc

But → biased estimation methods

iii- reparametrization

Impose data exactable restriction

Example: Prod Func: Under CES only are expl. vars

Consider the following aggregate export function for a small open economy

$$\ln \underbrace{EXP_t}_{\substack{\downarrow \\ \text{aggregate} \\ \text{real export}}} = \hat{b}_0 + \hat{b}_1 \underbrace{\ln RY_t}_{\substack{\downarrow \\ \text{real} \\ \text{domestic} \\ \text{output}}} + \hat{b}_2 \underbrace{RER_t}_{\substack{\downarrow \\ \text{real} \\ \text{affective} \\ \text{rate}}}$$

RWY = Real World Output

international trade theory postulates that $\hat{b}_1 > 0$, $\hat{b}_2 < 0$ and $\hat{b}_3 > 0$

Consider the following equations estimated using 54 observations

Model 1

$$\ln EXP_t = 3.0 + 1.07 \ln RY_t - 1.10 RER_t - 0.4 \ln RWY_t$$

(1.4) (0.15) (0.15) (0.4)

$$R^2_1 = 0.918 \quad SSR_1 = 0.670$$

Model 2

$$\ln \hat{EXP}_t = 3.01 + 1.05 \ln RY_t - 1.11 RER_t$$

(1.4) (0.15) (0.15)

$$R^2_2 = 0.917 \quad SSR_2 = 0.671$$

Model 3

$$\ln \left(\frac{\hat{EXP}_t}{RY_t} \right) = 3.01 - 1.13 RER_t$$

(0.54) (0.14)

$$R^2_3 = 0.627 \quad SSR_3 = 0.673$$

Model 4

$$\ln EXP_t - \ln RY_t + RER_t = 3.01 \quad SSR_4 = 0.700$$

(0.02)

1- interpret model 1. Is it a valid model to explain aggregate export? perform all necessary tests

1- convert the notation

$$\ln EXP_t = \beta_0 + \beta_1 \ln RY_t + \beta_2 RER_t + \beta_3 \ln RUY_t + U_t$$

2- individual significance test

$$t_{50-4}^{0.025} = 2.008$$

3- construct the null & alternative hypothesis and test whether we reject or fail to reject

β_1	$H_0: \beta_1 = 0$	β_2	$H_0: \beta_2 = 0$	β_3	$H_0: \beta_3 = 0$
	$H_1: \beta_1 \neq 0$		$H_1: \beta_2 \neq 0$		$H_1: \beta_3 \neq 0$

then

$$|t_{\beta_1}| = \left| \frac{1.07 - 0}{0.15} \right| = 7.13 > 2.008 \quad \text{reject } H_0$$

$$|t_{\beta_2}| = \left| \frac{1.10 - 0}{0.15} \right| = 7.33 > 2.008 \quad \text{reject } H_0$$

$$|t_3| = \left| \frac{0.01 - 0}{0.4} \right| = 0.1 < 2.008 \quad \text{fail to reject } H_0$$

β_1 & β_2 are statistically significant

β_3 is not statistically significant

4 - joint significance test

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$

H_1 : at least one differ from 0

5 - construct restricted & unrestricted model

$$\ln EXP_t = \beta_0 + \beta_1 \ln RY_t + \beta_2 RER_t + \beta_3 \ln RWY_t + u_t$$

(unrestricted model)

$$\ln EXP_t = \beta_0 + u_t \rightarrow \text{restricted model}$$

6 - use F statistic

$$F_{cal} = \frac{R_u^2 - R_r^2}{1 - R_u^2} \cdot \frac{df_u}{9}$$

$$\frac{0.918 - 0}{1 - 0.918} \times \frac{50}{3} = 186.6$$

$$F_{3,5}^{0.05} = 2.80$$

$$F_{cal} > F_{3,50}^{0.05}$$

reject null hypothesis. coefficients are jointly significant

ii - compare model 1 & 2

$$(1) \ln EXP_t = \beta_0 + \beta_1 \ln RY_t + \beta_2 RER_t + \beta_3 \ln RWY_t + u_t$$

$$(2) \ln EXP_t = \beta_0 + \beta_1 \ln RY_t + \beta_2 RER_t + v_t$$

null hypothesis $H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

$$F_{cal} = \frac{R_u^2 - R_r^2}{1 - R_u^2} \cdot \frac{df_u}{9}$$

$$\frac{0.818 - 0.917}{1 - 0.918} \times \frac{50}{1} = 0.0041$$

$$F_{1, 50}^{0.05} = 4.03$$

$$F_{cal} < F_{1, 50}^{0.05}$$

fail to reject null hypothesis. $\rightarrow$ Restriction is valid

Model 2 better than 1

ii) Consider model 2

a) test the hypothesis that the coefficient of $\ln RY_t$ unity

$$H_0: \beta_1 = 1$$

$$H_1: \beta_1 \neq 1$$

1- calculate t-statistic

$$t_{cal} = \frac{\hat{\beta}_1 - \beta_1}{S.e(\hat{\beta}_1)} = \frac{1.05 - 1}{0.15} = 0.33$$

$$t_{0.025, 51} = 2.007$$

$$|t_{cal}| < t$$

fail to reject null hypothesis

b) test the hypothesis that the coefficients of $\ln RY_t$ and $\ln RER_t$ are jointly 1 & -1

$$H_0: \beta_1 = 1 \quad \beta_2 = -1$$

H_1 : o.w.

under the null, we have model 4, so model 4 turns out to be the restricted model, and we have model 2 as an unrestricted model:

since dependent variables are different, we cannot use R^2 to compare models.

$$F_{cal} = \frac{SSR_r - SSR_u}{SSR_u} \cdot \frac{df_u}{9}$$

$$F_{cal} = \frac{0.700 - 0.671}{0.671} \cdot \frac{51}{2} = 1.102$$

$$F_{2,51}^{0.05} = 3.178$$

$F_{cal} < F$
fail to reject null hypothesis

iv) it is argued that model 4 is not a valid econometric specification since there are more than one left-hand side variable

Model 4 is a valid economic specification. There is no more than one dependent variable.

Dependent variable is $\ln EXP_t - \ln RY_t + RER_t$

v) is it possible to compute R^2 for model 4 by using the information provided. If no, explain why.

We can compute R^2 , however since there is no explanatory variable we end up with $R^2 = 0$