Chapter 8
Multicollinearity

Masao Ogaki

Department of Economics,
Ohio State University

November 17, 2007

Perfect versus Imperfect Multicollinearity

Perfect Multicollinearity

Perfect multicollinearity is a violation of Classical Assumption VI.

We consider two demand functions for food

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \epsilon_i \tag{1}$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{3i} + \epsilon_i \tag{2}$$

where Y_i is demand for food, X_{1i} is the price of food, X_{2i} is wage income,and X_{3i} is total income at time i.

Do you expect α_1 and β_1 to be positive or negative? How about α_2 and α_2?

Imagine that

$$X_{1i} = 5 - 2X_{2i} \tag{3}$$

without any error term for the data set by chance.

If we try to run regression (1) in Excel, we get an error message "Linst0 function returns error." This is because this example is constructed so that X_1 and X_2 are perfectly linearly related.

We have *perfect multicollinearity.*

If we substitute Equation (3) into Equation (1), we obtain

$$\begin{aligned} Y_i = & \quad \alpha_0 + \alpha_1(5 - 2X_{2i}) + \alpha_2 X_{2i} + \epsilon_i \\ &= (\alpha_0 + 5\alpha_1) + (-2\alpha_1 + \alpha_2)X_{2i} + \epsilon_i \\ &= \gamma_0 + \gamma_1 X_{1i} + \epsilon_i \end{aligned} \tag{4}$$

where $\gamma_0 = \alpha_0 + 5\alpha_1$ and $\gamma_1 = -2\alpha_1 + \alpha_2$.

Hence we can estimate γ_0 and γ_1 (How?), but we cannot obtain estimates for α_0, α_1, and α_2.

Imperfect Multicollinearity
When some of the explanatory variables have near perfect linear relationship, we have *imperfect multicollinearity*. Since perfect multicollinearity is rare, if we just say multicollinearity, it means near multicollinearity.

Imagine that when we run Regression (2), we obtain

$$
\begin{array}{rrll}
\hat{Y}_i = 140.73 & -3.295 & X_{1i} & - 0.4189\ X_{3i} \\
 & (0.9154) & & (0.4375) \\
t = & -3.599 & & -0.959 \\
 & N = 16, & R^2 = & 0.9667
\end{array}
$$

Exercise: Does the coefficient of the income variable have the expected sign? Is this coefficient statistically significant at the 5 % level?

Imagine that we regress X_3 onto X_1, and obtain

$$\begin{aligned} \hat{Y}_i = 4.217 \quad & -1.92 \quad X_i \\ & (0.56) \\ t = \quad & -3.43 \\ N = 16, R^2 = \quad & 0.9887 \end{aligned}$$

The high value of R^2 indicates that price and total income have imperfect multicollinearity.

Consequences of Multicollinearity

In cases of imperfect multicollinearity, we are likely to encounter one or more of the following consequences:

1. Large standard error of OLS estimators.
2. Insignificant t-values. Note that it is possible to have a high R^2.
3. OLS estimates become very sensitive to small changes in the data; that is, they tend to be unstable.
4. Wrong signs for regression coefficients.

Exercise: Suppose that all Classical Assumptions are satisfied for Regression (2). Are the OLS estimators BLUE when we have the imperfect multicollinearity problem? Are they efficient?

Remedies of Multicollinearity

If we have symptoms listed in the previous section, you are likely to have multicollinearity. What should we do?

Do Nothing: Multicollinearity is not a serious problem if the purpose of the study is to predict or forecast future values of the dependent variable. Even though we cannot precisely estimate individual regression coefficients,the predicted value may be stable even when estimates are unstable.

However, if the objective of the study is not only prediction but also reliable estimation of individual regression coefficients, then multicollinearity is a serious problem. In this case, we may try one or more of the following methods.

1. **Drop a Redundant Variable**
2. **Use Different Functional Forms**
3. **Use Prior Information about Some Parameters**
4. **Increase the Size of the Sample**