# MULTIKOLINEARITAS

EKONOMETRIKA 1

AL MUIZZUDDIN F., SE., ME. (2015)

FEB UB

# What Happens If the Regressors Are Correlated?

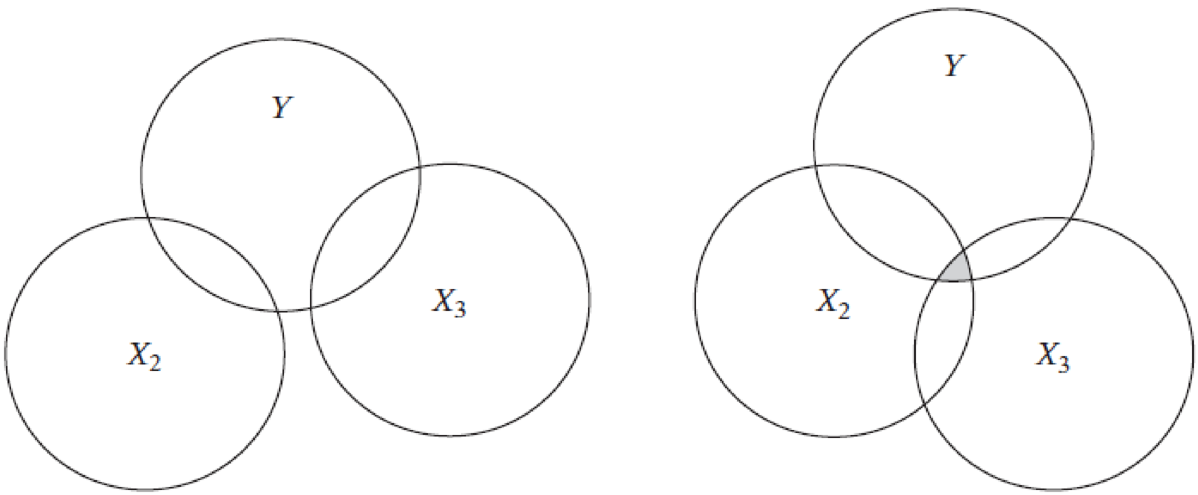
## In this chapter we take a critical look at this assumption by seeking answers to the following questions:

- 1. What is the nature of multicollinearity?
- 2. Is multicollinearity really a problem?
- 3. What are its practical consequences?
- 4. How does one detect it?
- 5. What remedial measures can be taken to alleviate the problem of multicollinearity?

# The Nature of Multicollinearity

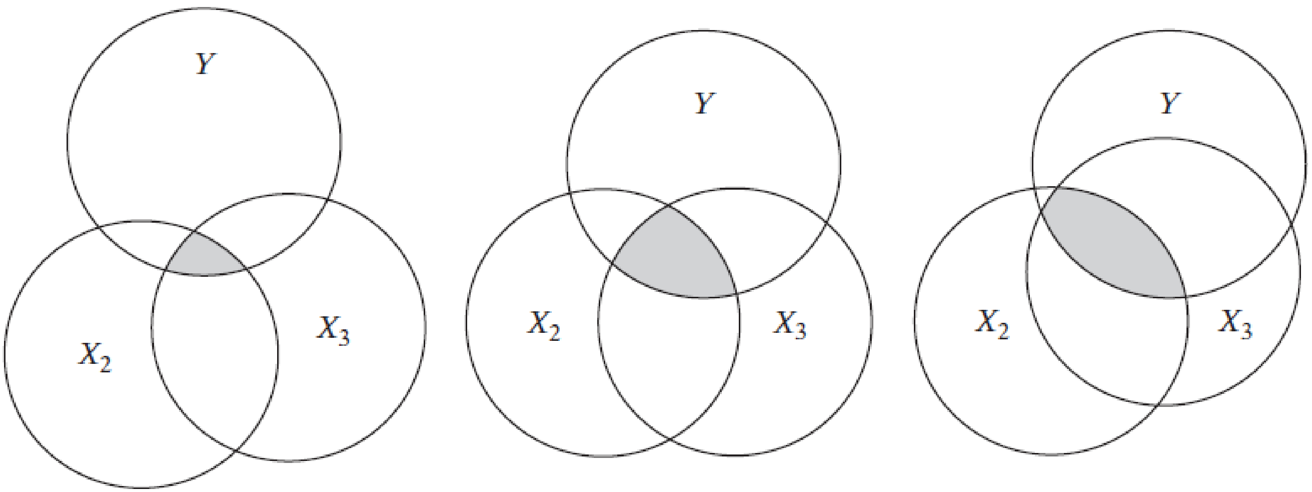
- The term *multicollinearity* is due to Ragnar Frisch.3 Originally it meant the existence of a "perfect," or exact, linear relationship among some or all explanatory variables of a regression model.

# The preceding algebraic approach to multicollinearity can be portrayed succinctly by the Ballentine




(*a*) No collinearity (*b*) Low collinearity

# Next




(*c*) Moderate collinearity (*d*) High collinearity (*e*) Very high collinearity

# Why does the classical linear regression model assume that there is no multicollinearity among the *X*'s?

- **If multicollinearity is perfect in the sense of Eq. (10.1.1), the regression coefficients of the *X* variables <u>are indeterminate</u> and their <u>standard errors are infinite</u>. If multicollinearity is less than perfect, as in Eq. (10.1.2), the regression coefficients, although determinate, possess <u>large standard errors</u> (in relation to the coefficients themselves), which means <u>the coefficients cannot be estimated</u> with great precision or accuracy.**

# There are several sources of multicollinearity

1. *There are several sources of multicollinearity.*
2. *Constraints on the model or in the population being sampled*
3. *Model specification*
4. *An overdetermined model.*

# Theoretical Consequences of Multicollinearity

- First, it is true that even in the case of near multicollinearity the OLS estimators are unbiased.
- Second, it is also true that collinearity does not destroy the property of minimum variance: In the class of all linear unbiased estimators, the OLS estimators have minimum variance; that is, they are efficient.
- Third, *multicollinearity is essentially a sample (regression) phenomenon* in the sense that, even if the $X$ variables are not linearly related in the population, they may be so related in the particular sample at hand: When we postulate the theoretical or population regression function (PRF), we believe that all the $X$ variables included in the model have a separate or independent influence on the dependent variable $Y$.

# Practical Consequences of Multicollinearity

- 1. Although BLUE, the OLS estimators have large variances and covariances, making precise estimation difficult.
- 2. Because of consequence 1, the confidence intervals tend to be much wider, leading to the acceptance of the "zero null hypothesis" (i.e., the true population coefficient is zero) more readily.
- 3. Also because of consequence 1, the *t* ratio of one or more coefficients tends to be statistically insignificant.
- 4. Although the *t* ratio of one or more coefficients is statistically insignificant, $R2$, the overall measure of goodness of fit, can be very high.
- 5. The OLS estimators and their standard errors can be sensitive to small changes in the data.

# An Illustrative Example

**EXAMPLE 10.1**

*Consumption Expenditure in Relation to Income and Wealth*

To illustrate the various points made thus far, let us consider the consumption–income example from the introduction. Table 10.5 contains hypothetical data on consumption, income, and wealth. If we assume that consumption expenditure is linearly related to income and wealth, then, from Table 10.5 we obtain the following regression:

$$
\begin{aligned}
\hat{Y}_i &= 24.7747 + 0.9415X_{2i} - 0.0424X_{3i} \\
&\quad (6.7525) \quad (0.8229) \quad (0.0807) \\
t &= (3.6690) \quad (1.1442) \quad (-0.5261) \\
R^2 &= 0.9635 \quad \bar{R}^2 = 0.9531 \quad \text{df} = 7
\end{aligned}
\qquad (10.6.1)
$$

**TABLE 10.5 Hypothetical Data on Consumption Expenditure $Y$, Income $X_2$, and Wealth $X_3$**

| $Y$, \$ | $X_2$, \$ | $X_3$, \$ |
|---|---|---|
| 70 | 80 | 810 |
| 65 | 100 | 1009 |
| 90 | 120 | 1273 |
| 95 | 140 | 1425 |
| 110 | 160 | 1633 |
| 115 | 180 | 1876 |
| 120 | 200 | 2052 |
| 140 | 220 | 2201 |
| 155 | 240 | 2435 |
| 150 | 260 | 2686 |

Regression (10.6.1) shows that income and wealth together explain about 96 percent of the variation in consumption expenditure, and yet neither of the slope coefficients is individually statistically significant. Moreover, not only is the wealth variable statistically insignificant but also it has the wrong sign. A priori, one would expect a positive relationship between consumption and wealth. Although $\hat{\beta}_2$ and $\hat{\beta}_3$ are individually statistically insignificant, if we test the hypothesis that $\beta_2 = \beta_3 = 0$ simultaneously, this hypothesis can be rejected, as Table 10.6 shows. Under the usual assumption we obtain

$$F = \frac{4282.7770}{46.3494} = 92.4019 \quad \text{(10.6.2)}$$

This $F$ value is obviously highly significant.

# Detection of Multicollinearity

- 1. **High *R*2 but few significant *t* ratios.**
- 2. **High pair-wise correlations among regressors.**
- 3. **Tolerance and variance inflation factor.**

# Remedial Measures

- **Do Nothing :** Multicollinearity is God's will, not a problem with OLS or statistical technique in general (Blanchard).
- **Rule-of-Thumb Procedures**
  - 1. **A priori information.**
  - 2. **Combining cross-sectional and time series data**
  - 3. **Dropping a variable(s) and specification bias.**
  - 4. **Transformation of variables.**
  - 5. **Additional or new data.**

# STOP