

*Essential  
Chapter 7*

240 PART TWO: REGRESSION ANALYSIS IN PRACTICE

Since the search for a theoretically correct model can be exasperating, in this chapter we considered several practical criteria that we should keep in mind in this search, such as (1) parsimony, (2) identifiability, (3) goodness of fit, (4) theoretical consistency, and (5) predictive power.

As Granger notes, "In the ultimate analysis, model building is probably both an art and a science. A sound knowledge of theoretical econometrics and the availability of an efficient computer program are not enough to ensure success."<sup>16</sup>

### KEY TERMS AND CONCEPTS

The key terms and concepts introduced in this chapter are:

Attributes of a good model

- a) Parsimony (principle of parsimony)
- b) Identifiability
- c) Goodness of fit
- d) Theoretical consistency
- e) Predictive power

Specification errors and model misspecification errors

- a) Core variables
- b) Peripheral variables
- c) Underfitting a model (omitted variable bias)
- d) Overfitting a model (inclusion of irrelevant variable bias)

e) Incorrect (wrong) functional form bias

f) Instrumental or proxy variables

Specification error tests

- a) Unnecessary variables (stepwise regression; data mining)
- b) Tests for omitted variables and incorrect functional forms
- c) MacKinnon-White-Davidson (MWD) test
- d) Ramsey's regression error specification (RESET) test

### QUESTIONS

- 7.1. What is meant by specification errors?
- 7.2. What are the reasons for the occurrence of specification errors?
- 7.3. What are the attributes of a "good" econometric model?
- 7.4. What are different types of specification errors? Can one or more of these errors occur simultaneously?
- 7.5. What are the consequences of omitting a relevant variable(s) from a model?
- 7.6. When we say that a variable is "relevant" or "irrelevant," what do we mean?
- 7.7. What are the consequences of including irrelevant variables in a model?
- 7.8. Omitting a relevant variable(s) from a model is more dangerous than including an irrelevant variable(s). Do you agree? Why or why not?

<sup>16</sup>See C. W. J. Granger (ed.), *Modelling Economic Time Series: Readings in Econometric Methodology*, Clarendon, Oxford, U.K., 1990, p. 2.

- 7.9. In looking for the simple Keynesian multiplier, you regress the GNP on investment and find that there is some relationship. Now, thinking that it cannot hurt much, you include the "irrelevant" variable "state and local taxes." To your surprise, the investment variable loses its significance. How can an irrelevant variable do this?
- 7.10. What would you do if you had to choose between a model that satisfies all statistical criteria but does not satisfy economic theory and a model that fits established economic theory but does not fit many statistical criteria?

**PROBLEMS**

- 7.11. Table 7-5, found on the textbook's Web site, gives data on the real gross product, labor input, and real capital input in the Taiwanese manufacturing sector for the years 1958 to 1972. Suppose the theoretically correct production function is of the Cobb-Douglas type, as follows:

$$\ln Y_t = B_1 + B_2 \ln X_{2t} + B_3 \ln X_{3t} + u_t$$

where  $\ln$  = the natural log.

- a. Given the data shown in Table 7-5, estimate the Cobb-Douglas production function for Taiwan for the sample period and interpret the results.
- b. Suppose capital data were not initially available and therefore someone estimated the following production function:

$$\ln Y_t = C_1 + C_2 \ln X_{2t} + v_t$$

where  $v$  = an error term. What kind of specification error is incurred in this case? What are the consequences? Illustrate with the data in Table 7-5.

- c. Now pretend that the data on labor input were not available initially and suppose you estimated the following model:

$$\ln Y_t = C_1 + C_2 \ln X_{3t} + w_t$$

where  $w$  = an error term. What are the consequences of this type of specification error? Illustrate with the data given in Table 7-5.

- 7.12. Consider the following models:  
 Model I:  $\text{Consumption}_i = B_1 + B_2 \text{income}_i + u_i$   
 Model II:  $\text{Consumption}_i = A_1 + A_2 \text{wealth}_i + v_i$
- a. How would you decide which of the models is the "true" model?
- b. Suppose you regress consumption on both income and wealth. How would this help you decide between the two models? Show the necessary details.
- 7.13. Refer to Equation 5.40 in Chapter 5, which discusses the regression-through-the-origin (i.e., zero-intercept) model. If there is in fact an intercept present in the model but you run it through the origin, what kind of specification error is committed? Document the consequences of this type of error with the data given in Table 2-13 (found on the textbook's Web site) in Chapter 2.