I.3 METHODOLOGY OF ECONOMETRICS

How do econometricians proceed in their analysis of an economic problem? That is, what is their methodology? Although there are several schools of thought on econometric methodology, we present here the **traditional** or **classical** methodology, which still dominates empirical research in economics and other social and behavioral sciences.[9]

Broadly speaking, traditional econometric methodology proceeds along the following lines:

1. Statement of theory or hypothesis.
2. Specification of the mathematical model of the theory
3. Specification of the statistical, or econometric, model
4. Obtaining the data
5. Estimation of the parameters of the econometric model
6. Hypothesis testing
7. Forecasting or prediction
8. Using the model for control or policy purposes.

To illustrate the preceding steps, let us consider the well-known Keynesian theory of consumption.

---

[8]Aris Spanos, *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data,* Cambridge University Press, United Kingdom, 1999, p. 21.

[9]For an enlightening, if advanced, discussion on econometric methodology, see David F. Hendry, *Dynamic Econometrics*, Oxford University Press, New York, 1995. See also Aris Spanos, *op. cit.*

1. Statement of Theory or Hypothesis

Keynes stated:

> The fundamental psychological law . . . is that men [women] are disposed, as a rule and on average, to increase their consumption as their income increases, but not as much as the increase in their income.[10]

In short, Keynes postulated that the **marginal propensity to consume (MPC),** the rate of change of consumption for a unit (say, a dollar) change in income, is greater than zero but less than 1.

2. Specification of the Mathematical Model of Consumption

Although Keynes postulated a positive relationship between consumption and income, he did not specify the precise form of the functional relationship between the two. For simplicity, a mathematical economist might suggest the following form of the Keynesian consumption function:

$$Y = \beta_1 + \beta_2 X \qquad 0 < \beta_2 < 1 \tag{I.3.1}$$

where Y = consumption expenditure and X = income, and where β_1 and β_2, known as the **parameters** of the model, are, respectively, the **intercept** and **slope** coefficients.

The slope coefficient β_2 measures the MPC. Geometrically, Eq. (I.3.1) is as shown in Figure I.1. This equation, which states that consumption is lin-

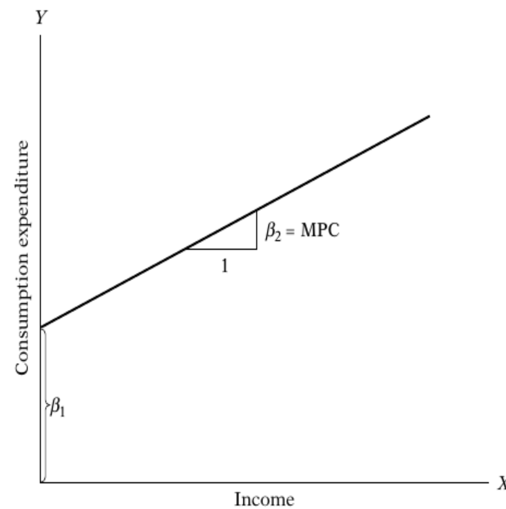

FIGURE I.1 Keynesian consumption function.

[10]John Maynard Keynes, *The General Theory of Employment, Interest and Money,* Harcourt Brace Jovanovich, New York, 1936, p. 96.

early related to income, is an example of a mathematical model of the relationship between consumption and income that is called the **consumption function** in economics. A model is simply a set of mathematical equations. If the model has only one equation, as in the preceding example, it is called a **single-equation model,** whereas if it has more than one equation, it is known as a **multiple-equation model** (the latter will be considered later in the book).

In Eq. (I.3.1) the variable appearing on the left side of the equality sign is called the **dependent variable** and the variable(s) on the right side are called the **independent,** or **explanatory,** variable(s). Thus, in the Keynesian consumption function, Eq. (I.3.1), consumption (expenditure) is the dependent variable and income is the explanatory variable.

3. Specification of the Econometric Model of Consumption

The purely mathematical model of the consumption function given in Eq. (I.3.1) is of limited interest to the econometrician, for it assumes that there is an *exact* or *deterministic* relationship between consumption and income. But relationships between economic variables are generally inexact. Thus, if we were to obtain data on consumption expenditure and disposable (i.e., aftertax) income of a sample of, say, 500 American families and plot these data on a graph paper with consumption expenditure on the vertical axis and disposable income on the horizontal axis, we would not expect all 500 observations to lie exactly on the straight line of Eq. (I.3.1) because, in addition to income, other variables affect consumption expenditure. For example, size of family, ages of the members in the family, family religion, etc., are likely to exert some influence on consumption.

To allow for the inexact relationships between economic variables, the econometrician would modify the deterministic consumption function (I.3.1) as follows:

$$Y = \beta_1 + \beta_2 X + u \qquad \textbf{(I.3.2)}$$

where u, known as the **disturbance,** or **error, term,** is a **random (stochastic) variable** that has well-defined probabilistic properties. The disturbance term u may well represent all those factors that affect consumption but are not taken into account explicitly.

Equation (I.3.2) is an example of an **econometric model.** More technically, it is an example of a **linear regression model,** which is the major concern of this book. The econometric consumption function hypothesizes that the dependent variable Y (consumption) is linearly related to the explanatory variable X (income) but that the relationship between the two is not exact; it is subject to individual variation.

The econometric model of the consumption function can be depicted as shown in Figure I.2.

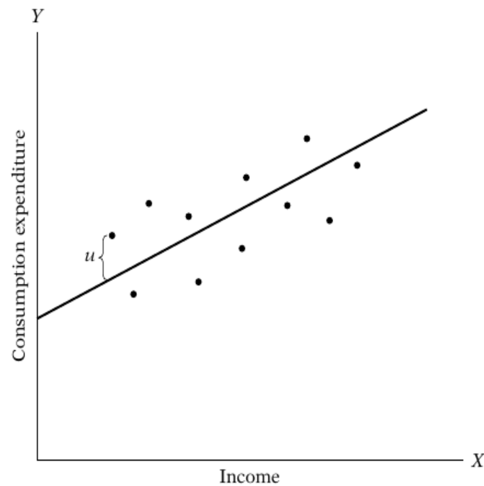

FIGURE I.2 Econometric model of the Keynesian consumption function.

4. Obtaining Data

To estimate the econometric model given in (I.3.2), that is, to obtain the numerical values of β_1 and β_2, we need data. Although we will have more to say about the crucial importance of data for economic analysis in the next chapter, for now let us look at the data given in Table I.1, which relate to

TABLE I.1 DATA ON *Y* (PERSONAL CONSUMPTION EXPENDITURE) AND *X* (GROSS DOMESTIC PRODUCT, 1982–1996), BOTH IN 1992 BILLIONS OF DOLLARS

Year	*Y*	*X*
1982	3081.5	4620.3
1983	3240.6	4803.7
1984	3407.6	5140.1
1985	3566.5	5323.5
1986	3708.7	5487.7
1987	3822.3	5649.5
1988	3972.7	5865.2
1989	4064.6	6062.0
1990	4132.2	6136.3
1991	4105.8	6079.4
1992	4219.8	6244.4
1993	4343.6	6389.6
1994	4486.0	6610.7
1995	4595.3	6742.1
1996	4714.1	6928.4

Source: Economic Report of the President, 1998, Table B–2, p. 282.

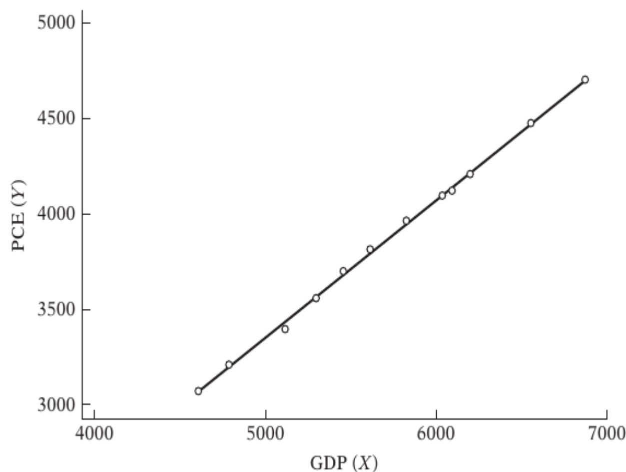

FIGURE I.3 Personal consumption expenditure (Y) in relation to GDP (X), 1982–1996, both in billions of 1992 dollars.

the U.S. economy for the period 1981–1996. The Y variable in this table is the *aggregate* (for the economy as a whole) personal consumption expenditure (PCE) and the X variable is gross domestic product (GDP), a measure of aggregate income, both measured in billions of 1992 dollars. Therefore, the data are in "real" terms; that is, they are measured in constant (1992) prices. The data are plotted in Figure I.3 (cf. Figure I.2). For the time being neglect the line drawn in the figure.

5. Estimation of the Econometric Model

Now that we have the data, our next task is to estimate the parameters of the consumption function. The numerical estimates of the parameters give empirical content to the consumption function. The actual mechanics of estimating the parameters will be discussed in Chapter 3. For now, note that the statistical technique of **regression analysis** is the main tool used to obtain the estimates. Using this technique and the data given in Table I.1, we obtain the following estimates of β_1 and β_2, namely, −184.08 and 0.7064. Thus, the estimated consumption function is:

$$\hat{Y} = -184.08 + 0.7064X_i \qquad \textbf{(I.3.3)}$$

The hat on the Y indicates that it is an estimate.[11] The estimated consumption function (i.e., regression line) is shown in Figure I.3.

[11]As a matter of convention, a hat over a variable or parameter indicates that it is an estimated value.

As Figure I.3 shows, the regression line fits the data quite well in that the data points are very close to the regression line. From this figure we see that for the period 1982–1996 the slope coefficient (i.e., the **MPC**) was about 0.70, suggesting that for the sample period an increase in real income of 1 dollar led, *on average*, to an increase of about 70 cents in real consumption expenditure.[12] We say *on average* because the relationship between consumption and income is inexact; as is clear from Figure I.3; not all the data points lie exactly on the regression line. In simple terms we can say that, according to our data, the *average*, or *mean*, consumption expenditure went up by about 70 cents for a dollar's increase in real income.

6. Hypothesis Testing

Assuming that the fitted model is a reasonably good approximation of reality, we have to develop suitable criteria to find out whether the estimates obtained in, say, Eq. (I.3.3) are in accord with the expectations of the theory that is being tested. According to "positive" economists like Milton Friedman, a theory or hypothesis that is not verifiable by appeal to empirical evidence may not be admissible as a part of scientific enquiry.[13]

As noted earlier, Keynes expected the MPC to be positive but less than 1. In our example we found the MPC to be about 0.70. But before we accept this finding as confirmation of Keynesian consumption theory, we must enquire whether this estimate is sufficiently below unity to convince us that this is not a chance occurrence or peculiarity of the particular data we have used. In other words, is 0.70 *statistically less than 1?* If it is, it may support Keynes' theory.

Such confirmation or refutation of economic theories on the basis of sample evidence is based on a branch of statistical theory known as **statistical inference (hypothesis testing).** Throughout this book we shall see how this inference process is actually conducted.

7. Forecasting or Prediction

If the chosen model does not refute the hypothesis or theory under consideration, we may use it to predict the future value(s) of the dependent, or **forecast, variable** Y on the basis of known or expected future value(s) of the explanatory, or **predictor, variable** X.

To illustrate, suppose we want to predict the mean consumption expenditure for 1997. The GDP value for 1997 was 7269.8 billion dollars.[14] Putting

---

[12]Do not worry now about how these values were obtained. As we show in Chap. 3, the statistical method of **least squares** has produced these estimates. Also, for now do not worry about the negative value of the intercept.

[13]See Milton Friedman, "The Methodology of Positive Economics," *Essays in Positive Economics,* University of Chicago Press, Chicago, 1953.

[14]Data on PCE and GDP were available for 1997 but we purposely left them out to illustrate the topic discussed in this section. As we will discuss in subsequent chapters, it is a good idea to save a portion of the data to find out how well the fitted model predicts the out-of-sample observations.

this GDP figure on the right-hand side of (I.3.3), we obtain:

$$\begin{aligned}\hat{Y}_{1997} &= -184.0779 + 0.7064\,(7269.8)\\ &= 4951.3167\end{aligned} \qquad \textbf{(I.3.4)}$$

or about 4951 billion dollars. Thus, given the value of the GDP, the mean, or average, forecast consumption expenditure is about 4951 billion dollars. The actual value of the consumption expenditure reported in 1997 was 4913.5 billion dollars. The estimated model (I.3.3) thus **overpredicted** the actual consumption expenditure by about 37.82 billion dollars. We could say the **forecast error** is about 37.82 billion dollars, which is about 0.76 percent of the actual GDP value for 1997. When we fully discuss the linear regression model in subsequent chapters, we will try to find out if such an error is "small" or "large." But what is important for now is to note that such forecast errors are inevitable given the statistical nature of our analysis.

There is another use of the estimated model (I.3.3). Suppose the President decides to propose a reduction in the income tax. What will be the effect of such a policy on income and thereby on consumption expenditure and ultimately on employment?

Suppose that, as a result of the proposed policy change, investment expenditure increases. What will be the effect on the economy? As macroeconomic theory shows, the change in income following, say, a dollar's worth of change in investment expenditure is given by the **income multiplier** ***M***, which is defined as

$$M = \frac{1}{1 - \text{MPC}} \qquad \textbf{(I.3.5)}$$

If we use the MPC of 0.70 obtained in (I.3.3), this multiplier becomes about $M = 3.33$. That is, an increase (decrease) of a dollar in investment will *eventually* lead to more than a threefold increase (decrease) in income; note that it takes time for the multiplier to work.

The critical value in this computation is MPC, for the multiplier depends on it. And this estimate of the MPC can be obtained from regression models such as (I.3.3). Thus, a quantitative estimate of MPC provides valuable information for policy purposes. Knowing MPC, one can predict the future course of income, consumption expenditure, and employment following a change in the government's fiscal policies.

8. Use of the Model for Control or Policy Purposes

Suppose we have the estimated consumption function given in (I.3.3). Suppose further the government believes that consumer expenditure of about 4900 (billions of 1992 dollars) will keep the unemployment rate at its

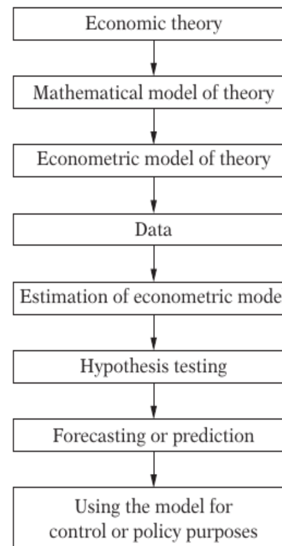

FIGURE I.4 Anatomy of econometric modeling.

current level of about 4.2 percent (early 2000). What level of income will guarantee the target amount of consumption expenditure?

If the regression results given in (I.3.3) seem reasonable, simple arithmetic will show that

$$4900 = -184.0779 + 0.7064X \tag{I.3.6}$$

which gives $X = 7197$, approximately. That is, an income level of about 7197 (billion) dollars, given an MPC of about 0.70, will produce an expenditure of about 4900 billion dollars.

As these calculations suggest, an estimated model may be used for control, or policy, purposes. By appropriate fiscal and monetary policy mix, the government can manipulate the **control variable *X*** to produce the desired level of the **target variable *Y*.**

Figure I.4 summarizes the anatomy of classical econometric modeling.

Choosing among Competing Models

When a governmental agency (e.g., the U.S. Department of Commerce) collects economic data, such as that shown in Table I.1, it does not necessarily have any economic theory in mind. How then does one know that the data really support the Keynesian theory of consumption? Is it because the Keynesian consumption function (i.e., the regression line) shown in Figure I.3 is extremely close to the actual data points? Is it possible that an-

other consumption model (theory) might equally fit the data as well? For example, Milton Friedman has developed a model of consumption, called the *permanent income hypothesis*.[15] Robert Hall has also developed a model of consumption, called the *life-cycle permanent income hypothesis*.[16] Could one or both of these models also fit the data in Table I.1?

In short, the question facing a researcher in practice is how to choose among competing hypotheses or models of a given phenomenon, such as the consumption–income relationship. As Miller contends:

> No encounter with data is step towards genuine confirmation unless the hypothesis does a better job of coping with the data than some natural rival. . . . What strengthens a hypothesis, here, is a victory that is, at the same time, a defeat for a plausible rival.[17]

How then does one choose among competing models or hypotheses? Here the advice given by Clive Granger is worth keeping in mind:[18]

> I would like to suggest that in the future, when you are presented with a new piece of theory or empirical model, you ask these questions:
>
> **(i)** What purpose does it have? What economic decisions does it help with? and;
> **(ii)** Is there any evidence being presented that allows me to evaluate its quality compared to alternative theories or models?
>
> I think attention to such questions will strengthen economic research and discussion.

As we progress through this book, we will come across several competing hypotheses trying to explain various economic phenomena. For example, students of economics are familiar with the concept of the production function, which is basically a relationship between output and inputs (say, capital and labor). In the literature, two of the best known are the *Cobb–Douglas* and the *constant elasticity of substitution* production functions. Given the data on output and inputs, we will have to find out which of the two production functions, if any, fits the data well.

The eight-step classical econometric methodology discussed above is neutral in the sense that it can be used to test any of these rival hypotheses.

Is it possible to develop a methodology that is comprehensive enough to include competing hypotheses? This is an involved and controversial topic.

[15]Milton Friedman, *A Theory of Consumption Function,* Princeton University Press, Princeton, N.J., 1957.

[16]R. Hall, "Stochastic Implications of the Life Cycle Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy,* 1978, vol. 86, pp. 971–987.

[17]R. W. Miller, *Fact and Method: Explanation, Confirmation, and Reality in the Natural and Social Sciences,* Princeton University Press, Princeton, N.J., 1978, p. 176.

[18]Clive W. J. Granger, *Empirical Modeling in Economics,* Cambridge University Press, U.K., 1999, p. 58.