CHAPTER ONE

INTRODUCTION: TYPES OF MOVEMENTS

ELEMENTARY MOVEMENTS

THE objective of economic dynamics is to describe and explain the fluctuations in economic magnitudes. In this first part of our study we shall limit ourselves to description. This description can be assisted greatly by the use of diagrams in which the movements, or changes, of the various economic magnitudes are represented by certain geometric figures. For a good understanding of the various types of movements it is necessary to have a minimum general knowledge of the geometric properties of these diagrams quite apart from the economic phenomena which they represent. By way of introduction we shall start out with these geometric characteristics.

Before we can study the relationship between the movements of one magnitude and the movements of one or more other magnitudes, it will be necessary to give a description of the movements of individual economic magnitudes. It is in the nature of the object of our study that these magnitudes vary; they are "variable magnitudes," or, briefly, "variables." The movements of these variables are often complicated. There is considerable advantage in separating the movements into components of a simpler character, which we may call "elementary movements." These we shall study first.

a) Systematic and random movements

A distinction has to be made between systematic and random movements. In the case of systematic movements the numbers representing successive magnitudes of one variable follow each other according to a design. Any such design is, however, absent from the magnitudes of successive observations of random variables. Even in the very simple case in which the variable measured can have only two values (0 and 1) a distinction between systematic and random movements can be made. The following two series, for instance, represent systematic movements:

0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, etc.
0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, etc.

There is obviously an infinite number of systematic movements; more and more complicated systems of succession of the different values can be devised. As has been mentioned, in a random movement all design is absent in the way in which the successive values are arranged. Such a random movement will occur, for instance, when one tosses a coin and marks 1 for heads and 0 for tails. The results of an infinite repetition of this game will show random movements. A variable showing such movements is sometimes called a "random variable." A "normally distributed random variable" satisfies two conditions, namely, (1) there is no design in the succession of its values and (2) the various values satisfy the frequency distribution of the Gaussian law. The distribution under the Gaussian law, also called the law of normal distribution or of normal errors, may be expressed in a mathematical formula which is represented diagrammatically in the shape of a bell. In a normal distribution all possible values will occur—not only, as in the example just quoted, the values 0 and 1. In a normal distribution there will be a relatively large number of small deviations from the average and relatively few large deviations from the average. A normally distributed random variable will be obtained in more complicated games of chance in which the number of possible values is large, or, strictly speaking, in which the number of possible values is infinite. If, for instance, one were to toss not one coin but a hundred and were each time to count as the result of one game the number of heads that turned up, the result would approximate very closely the movement of a normally distributed random variable. In general, a variable will be a normally distributed random variable if its values are the sum of the values of a large number of random variables that are independent of each other. In nontechnical language, the variable will appear as a normally distributed random variable if its fluctuations are due to a great number of independent small causes. The larger the number of causes, the more will the series of successive values approximate the movement of a normally distributed random variable.

b) Monotonic and periodic movements

There exist a great many different types of systematic movements. Within the scope of this book there is no need for a detailed mathematical treatment of them. A variety of functions and curves is discussed by textbooks in analysis and analytical geometry. One distinction, however, is of great importance, namely, that between monotonic and periodic movements. A monotonic movement is one which never reverses its direction. It may be either monotonically increasing or monotonically decreasing. In the first case, each successive value is greater than, or in the limiting case equal to, the previous value; in the second case, each successive value is smaller than, or in a limiting case equal to, the preceding value. A periodic movement, on the other hand, repeats itself exactly after a certain lapse of time. This lapse of time is called its "period." We also include in the category of periodic movements nonmonotonic movements which repeat themselves after a certain period in an enlarged or reduced form, in a certain proportion which we may indicate by a. Such movements will be called damped and antidamped movements, respectively. The following would be an example of a purely periodic movement:

$$4, 8, 12, 4, 8, 12, 4, 8, 12, \text{etc.}$$

An example of a damped periodic movement would be the following:

$$4, 8, 12, 2, 4, 6, 1, 2, 3, \text{etc.}$$

In both cases the period is three units of time; in the second case the movement repeats itself on a reduced scale with $a = \frac{1}{2}$.

We shall have to analyze some special cases of these two types of movements in somewhat greater detail.

The simplest example of a monotonic movement is a straight line. This line may either rise or fall, with the horizontal line as the intermediate case.

For a straight line the difference in level of two successive time units (years, quarters, or months) is always the same. This difference is called the "slope," or the "rate of increase." In accordance with mathematical usage we shall use the term "increase" also for declining lines, that is, for negative increases. A straight line has a constant rate of increase.

A second important example of monotonic movement is the exponential curve. A mathematical property of an exponential

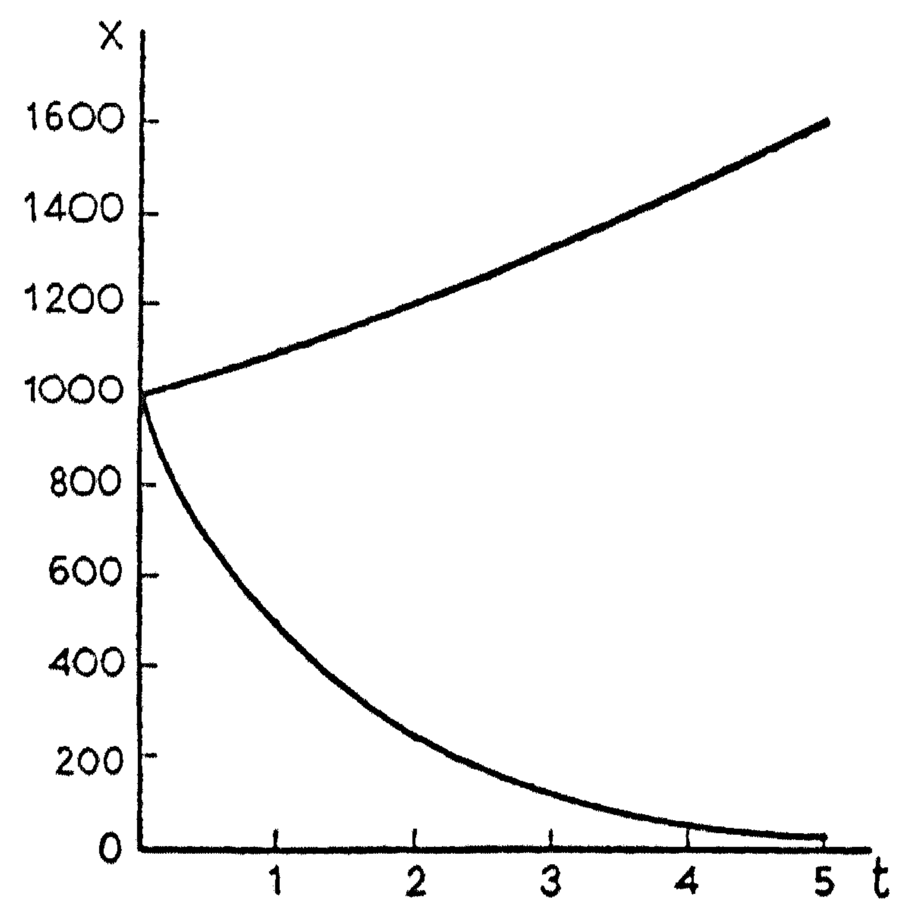

FIG. 1.—Exponential curves

curve is that its height at a series of successive equidistant points is indicated by a geometric series. That is to say that, in order to obtain a series of successive values, each at the same time distance from the preceding one, the height of the curve at each step has to be multiplied by a constant to obtain the next following step. Figure 1 shows two exponential curves. The first exponential curve is increasing. It is represented by the following figures:

Time t =	0	1	2	3	4	5	...
Height x =	1,000	1,100	1,210	1,331	1,464.1	1,610.51	...

The following series represents a declining exponential curve:

Time t =	0	1	2	3	4	5	...
Height x =	1,000	500	250	125	62.5	31.25	...

It will be noted that the slope, or rate of increase, of the exponential curve is not constant. It is proportional to the height attained by the curve. This may also be expressed by saying that the relative rate of increase is constant.

There are a number of other important monotonic movements. They cannot be treated without mathematical formulas. We mention here the parabolas of different degrees, represented by formulas of the form $x = at^n$, where n is the degree. These

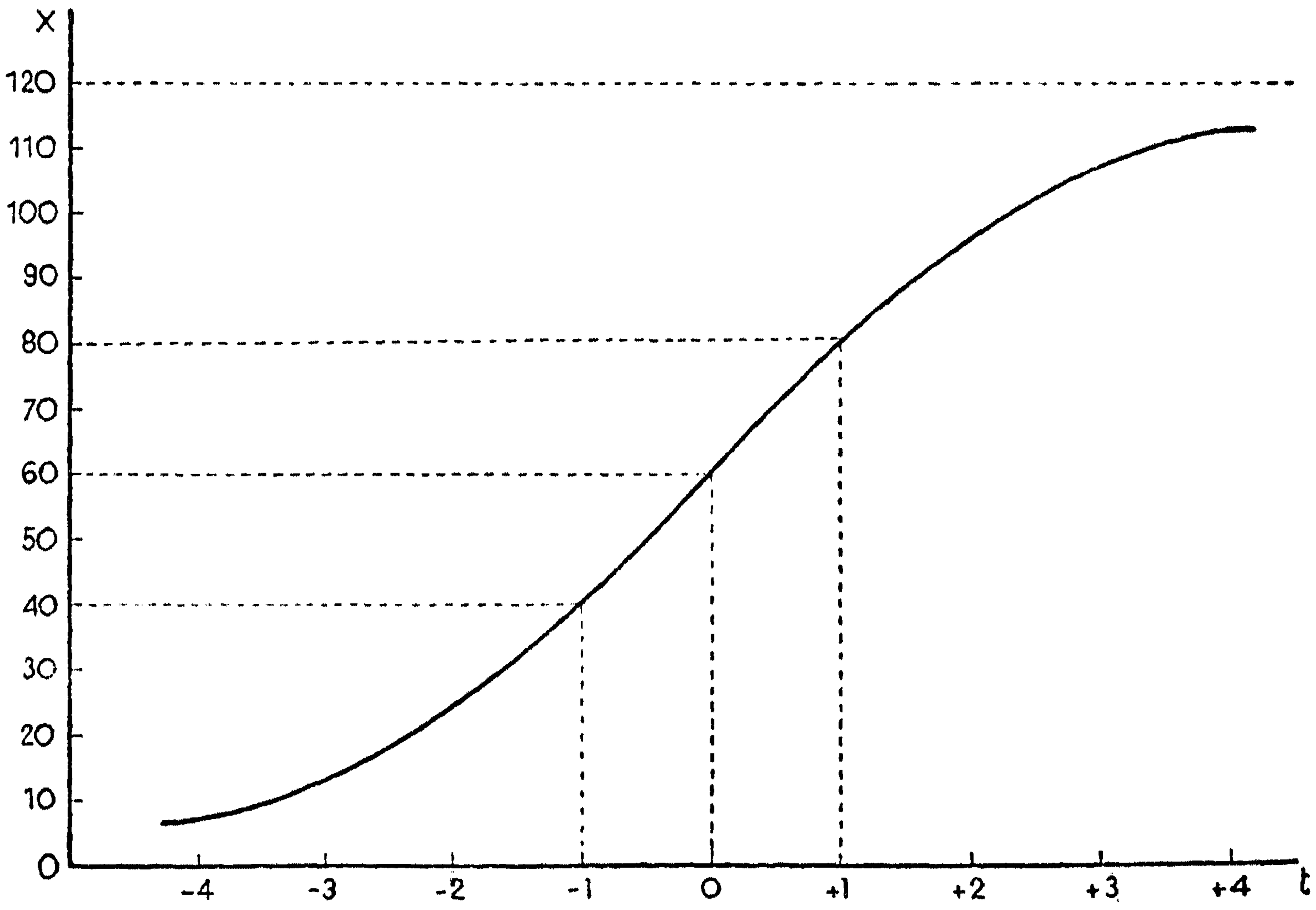

FIG. 2.—Logistic or growth curve ($a = 1$, $b = 2$, $k = 120$)

curves are monotonic for all positive values of t. Another monotonic curve of great importance is the logistic, or growth, curve: $x = k/(a + b^{-t})$. An example of the logistic curve is shown in Figure 2. A logistic curve will be described, for example, by the movement of a population, whose rate of growth per unit of time is proportional to (a) the size of the population already attained and (b) the size of the population for which, in addition to the population already present, means of subsistence are still available. The total means of subsistence are assumed to be such that a population of constant size can be maintained.

We count among the monotonic movements also certain movements that cannot be defined by one mathematical curve for the entire period of time but that are defined, for instance, first, by a rising straight line and, after a certain period, by a horizontal straight line.

A very simple type of periodic movement is the sine curve shown in Figure 3. As this diagram shows, the sine curve may be described as the distance from the horizontal axis of a point which moves at constant speed around the circumference of a circle. This distance is measured vertically while time is measured horizontally. The distance *AB* is called the period; the distances *CD* and *EF* are also equal to the period. Half the distance *AG* is called the "amplitude."

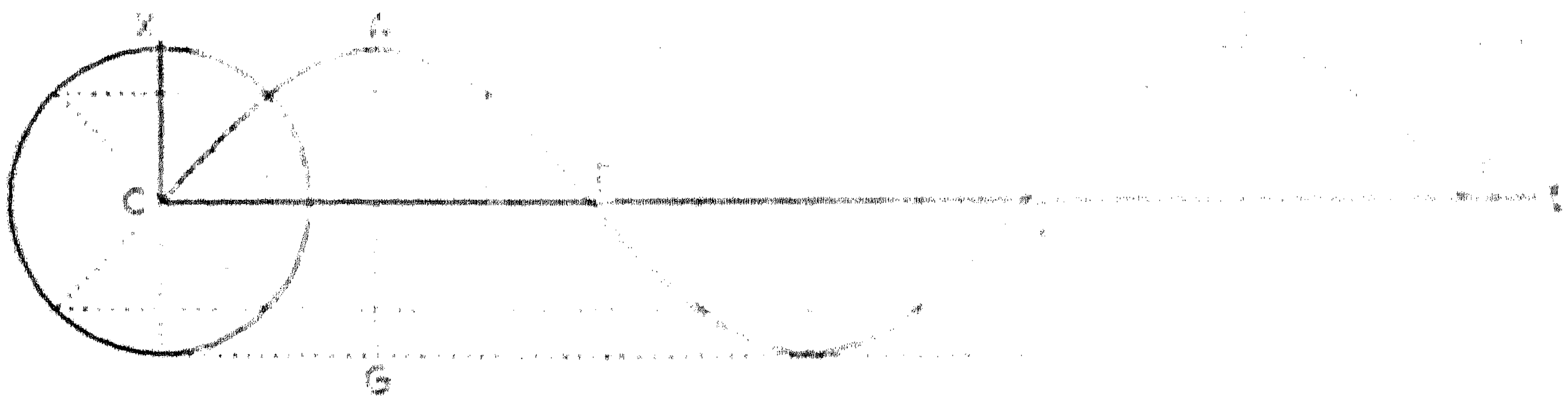

FIG. 3.—Sine curve (*C*, *A*, *E*, *D*, *B*, *F* . . .) indicating the distance from the horizontal axis of a point which moves around a circle with constant speed.

We shall often meet movements that are not exactly but approximately periodic in character. Such movements will show fluctuations repeating themselves approximately after roughly the same period. We shall sometimes refer to such movements as "fluctuating movements," or "fluctuations," or "waves." We may also call them "quasi-periodic movements." It should be noted here that random movements are also quasi-periodic. It may be proved that the quasi-period of these movements is three units of time, that is to say, that, on the average, one of every three values of such a series will be a peak and one a trough. The unit of time of such a series must be taken to be the distance between two successive independent observations.

c) *Damped, undamped, and antidamped movements*

As mentioned above, we include among the periodic movements those nonmonotonic movements that repeat themselves on an increasing or decreasing scale. We call periodic movements "damped" if each successive fluctuation is on a smaller scale than the preceding one, that is to say, if the factor a mentioned earlier is smaller than 1. The rate of damping is indicated by the fraction $1/a$. Movements which show antidamping are sometimes called "explosive movements." The purely periodic movement is undamped.

The same distinction can be made for exponential movements. If in an exponential movement the ratio between two successive values is smaller than 1, the movement is called damped; if the ratio is greater than 1, the movement is called antidamped, or explosive.

THE COMPONENTS OF COMPOSITE MOVEMENTS

The movements shown by most economic phenomena are much more complicated than any one of the simple movements we have discussed. First, any actual series will at most be an approximation of any one of the elementary movements considered above. Purely periodic movements, for instance, are rarely met in reality; in successive periods the shape of the movements repeats itself only approximately; often the period which one can distinguish is by no means constant. Second, actual series often represent combinations of elementary movements. We are using the mathematically rather vague expression "combinations" to cover not only the sum of elementary movements but also their product and in some instances even more complicated combinations. Whatever the mathematical nature of these combinations, all have in common the attribute that the actual movement can be constructed out of elementary movements.

Statisticians have adopted very generally the practice of separating the movements of economic variables into various elementary components before studying these components in further detail. In many respects this is a useful procedure; but a number of objections, to which we will refer below, can be made against it.

Normally, a series is decomposed into four components, as follows: (a) the trend component, which indicates the general tendency of the movement, and which usually is represented by a monotonic movement; (*b*) the cyclical component, consisting of fluctuations with a period of between three and eleven years; (*c*) the seasonal component, consisting of fluctuations with a period of one year, attributable to fluctuations of the natural and conventional seasons during the year; and (*d*) the random component, which covers both nonrecurrent changes, such as sudden changes in the level due to a "trend break," and fluctuations due to a large number of random causes, these latter fluctuations being usually of very short duration, e.g., those with a quasi-period of three months.

It is not advisable to apply this procedure mechanically, since there are many exceptions to the rule that these four components can be found in every economic series. We shall first mention a number of these exceptions and then treat the whole subject in a systematic manner.

Some fluctuations have a period in excess of eleven years. Among these should be counted the so-called "long cycles" and the fluctuations in certain individual markets. Depending on the subject and the period under consideration, such movements are classed under (*a*) or under (*b*); hence, a movement of type (*a*) cannot always be represented by a monotonic movement. There are, further, certain fluctuations which by their nature are equivalent to seasonal fluctuations but which have a shorter period, for instance, three months, a month, a week, or a day. On the other hand, random movements sometimes show quasi-periods that are considerably longer than three months. The fluctuations in crops, which must certainly to a large extent be considered as random, have a time interval of a year between two successive independent observations and hence may produce quasi-periods of three years. For these reasons it is often difficult to isolate components (*c*) and (*d*), or even (*b*) and (*d*), at all accurately.

On closer scrutiny, a variety of objectives may be discerned in the standard procedure. The first and clearest objective is to isolate movements with different periods. As a rule, all fluctuations with a period in excess of eleven years are comprised in the trend component. The monotonic movements of the trend component may be considered as parts of very long cycles or even of movements with an infinitely long period. All movements with a period of between one and eleven years are considered to be part of the cyclical component; those with a period of one year represent the seasonal component; and those with a period of less than one year, the random component. As mentioned above, however, it is sometimes difficult to classify the actual movements in accordance with this scheme.

A second objective in the standard procedure is the separation of systematic and random movements, discussed earlier.

The third objective would appear to be to obtain a classification according to the causes of the movements. The separation of seasonal movements and trend movements would reflect this objective. The trend movement, for instance, may often be ascribed to the very slow movements of such data as the size of the population, technical knowledge, etc.

A fully satisfactory decomposition of an economic variable into its various elements can be achieved only on the basis of a complete theory of economic movements. At this stage in our analysis, therefore, we can consider this decomposition only as a provisional tool. After having dealt with economic theory in the second part of this study, we shall in certain simple cases be able to give very definite directions with respect to the separation of the various elements (see chap. xvi). It will be shown that the systematic and random components are often combined in a very intricate fashion which renders it logically impossible to separate them in any simple way.

Here, as in the case of the separation of movements according to causes, a clear distinction between direct and indirect causes will have to be made. An example may make this clear. If fluctuations in the quantity demanded of a particular commodity are due to (*a*) changes in income and (*b*) changes in price, while the changes in price are due to (*c*) changes in the price of raw materials, then we call (*c*) an indirect cause of changes in the quantity demanded. On the other hand, (*a*) and (*b*) are considered direct causes.

It is often useful to isolate the components on the basis of the various direct causes; in such cases, however, indirect causes should be kept clearly separated from direct causes. The separation according to causes may not at all agree with the separation according to periods: one direct cause may be the origin of movements of different periods; one and the same period can occur in movements of different direct causes. In some instances the separation by causes and by periods may coincide. This will often be the case for seasonal movements that have both a special cause and a special period which does not coincide with the periods of movements due to other causes.

THE RELATIONSHIP BETWEEN THE MOVEMENTS OF TWO SERIES

Having discussed the various types of movements and the combination of these types of movements into one series, we shall now deal with the relationship that may exist between the movements of two series. These movements may be simple or composite.

We refer purposely to the relationship that *may* exist between the movements of two series. There need not be any relationship between them. The movements may be entirely independent. In reality there would usually be a certain degree of relationship. The examples treated below are the ideal case of a perfect relationship which can only be approximated in reality. Two series of figures show an exact relationship (or, in mathematical language, a functional relationship) if for any given value of the one series there is always a precisely defined value of the other series. Often no such precise relationship is present, but instead any given value of the one series is always found to be accompanied by approximately the same value of the other series. In statistics such a relationship is called a "stochastic relationship." In the following pages we shall discuss some of the most common functional relationships.

To indicate these relationships, we may make use of two types of diagrams: one shows the movement of the two series in time, preferably using the same time scale for both series; the other shows both series on a "scatter diagram." Such a diagram consists of a number of points in a rectangular system of co-ordinates. Each point has two co-ordinates, of which one, the x-co-ordinate, represents a value of the one series; and the other, the y-co-ordinate, the corresponding value of the other series. In such a diagram there are, therefore, as many points as there are values in each series. The time sequence of points cannot be traced in the diagram unless each point is labeled according to the time period to which it refers.

The simplest relationship between two series is that of equality. In that case the time diagrams of the two series cover each other in every point, if one adopts the same vertical scale for both series. In the scatter diagram, all points will lie on a straight line running through the origin at an angle of 45 degrees to both axes.

The next simplest relationship between two series is that of proportionality. In this case the time diagrams do not coincide, if the same vertical scale is used for both series. They can, however, be made to coincide by a special choice of the units for the vertical scale. If all figures of series Y are five times as large as those of series X, then the choice of the unit for Y at one-fifth of the unit for X will make the two series coincide. The scatter diagram still shows points on a straight line through the origin, but the slope of this line is now different. In this particular case the slope will be equal to a ratio of 5 : 1 (See Fig. 4).

One stage more complicated is the general linear relationship. In this relationship there is proportionality between the changes of the series but not between the absolute values of them; as a consequence, the ratio of the changes in the series is different from the ratio of their averages. The following two series give an example of this relationship:

X : 10, 12, 11, 14, 12, 13 (average 12), and
Y : 20, 22, 21, 24, 22, 23 (average 22).

Here the changes of the two series are equal, but the ratio of the averages is $\frac{22}{12}$, not $= 1$.

A second example is shown in Figure 5. The figure used in this diagram are as follows:

X : 10, 12, 11, 14, 12, 13 (average 12), and
Y : 20, 26, 23, 32, 26, 29 (average 26),

in which each change in Y is three times as large as the corresponding change in X, whereas the ratio of the averages of the

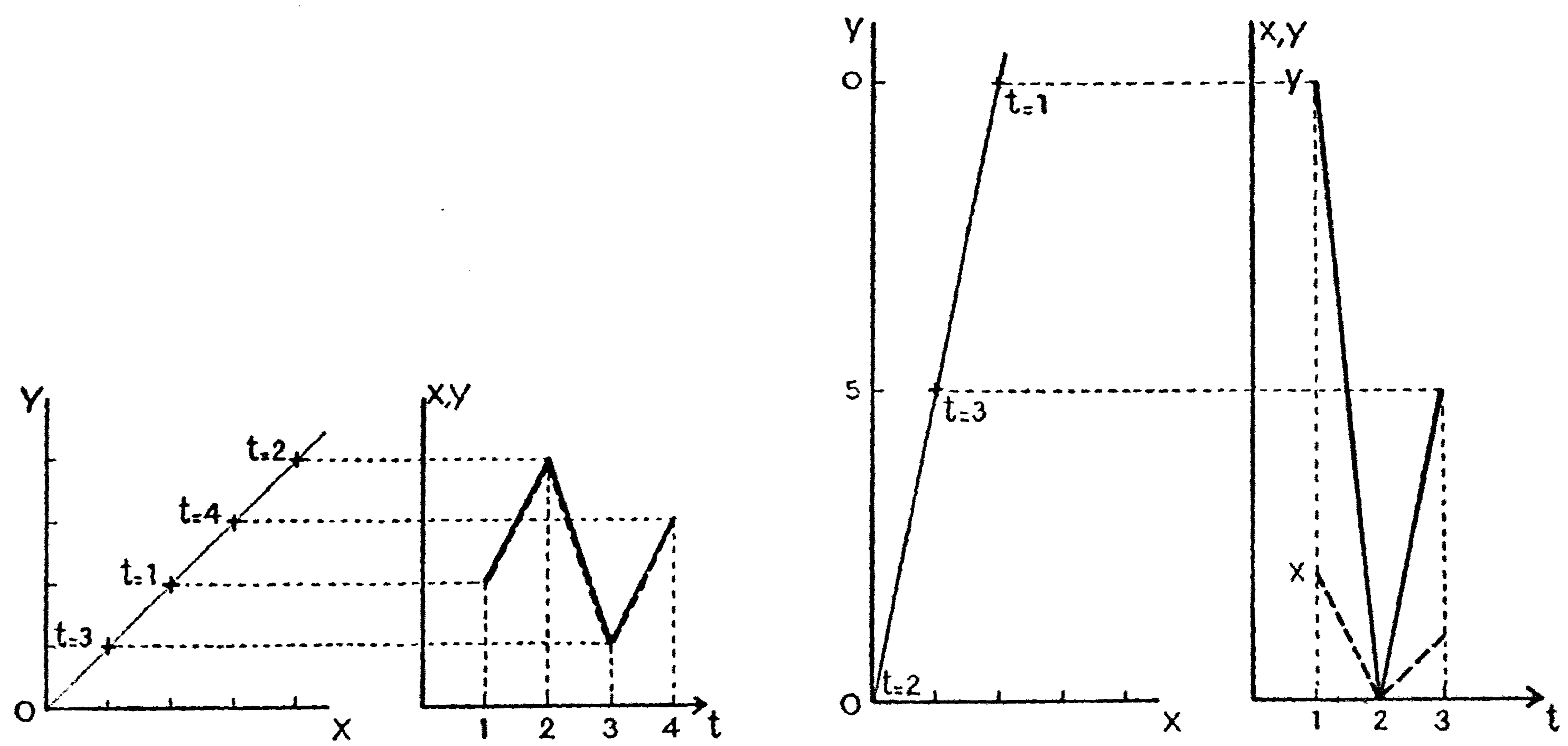

FIG. 4.—Equality (1st and 2d diagrams) and proportionality (3d and 4th diagrams) between X and Y. The first and third diagrams are scatter diagrams; the other two are time series.

two series is $26 \div 12$, obviously different from 3. If one wanted to make the graphs of these two time series coincide, it would not be sufficient to select different scales for them; it would also be necessary to draw the lines at different levels. In our first example it would be necessary to set the zero point for the Y series at the point -10 on the scale for X; in the second example it would be necessary to make the scale for Y one-third as great as the scale for X and, in addition, to set the zero point of the Y scale at the point $3\frac{1}{3}$ on the X scale.

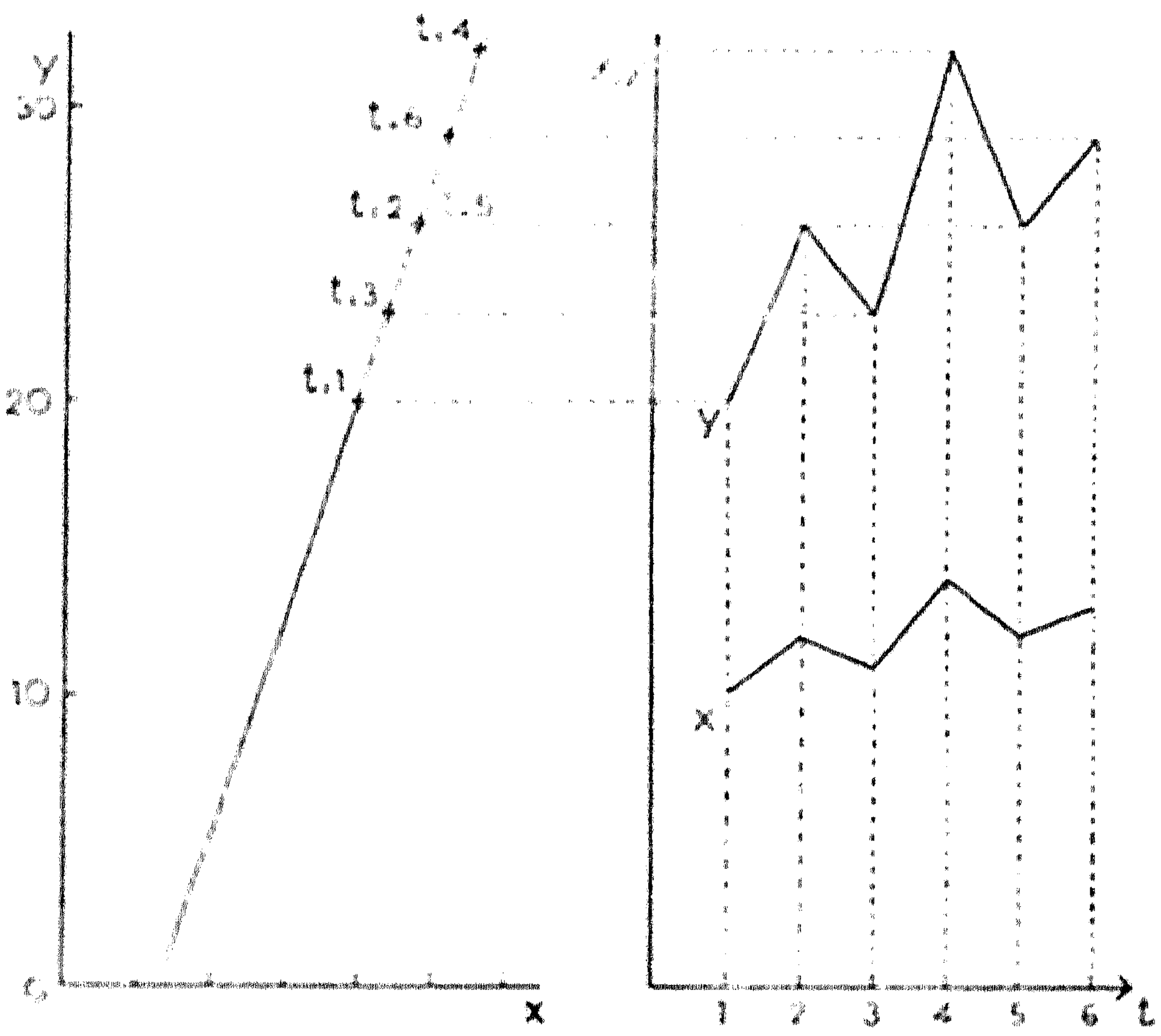

FIG. 5.—General linear relationship between X and Y

If two variables are related by a general linear relationship, it is sometimes said that their movement is parallel and also that there is a linear correlation between them. To distinguish this case from the one in which the relationship is only approximate in the stochastic sense, it is said that the correlation is perfect.

It will readily be realized that, depending on the absolute magnitudes of the fluctuations of the series and on their averages, one series may have larger absolute fluctuations while the other has larger relative fluctuations.

Equality, proportionality, and a general linear relationship may occur with a negative sign. In such instances the movements of the two series are in opposite directions.

The three forms of functional relationship discussed can all be represented by a straight line in a scatter diagram. We have discussed them in some detail because it will be found that they occur very frequently. There is, further, an infinite number of curvilinear functions that can be represented by other than straight lines in a scatter diagram. Curvilinear relationships can have many different shapes, depending on the nature of the relationship between the two variables. One example is given in Figure 6, which shows the relationship between the price, p, of a

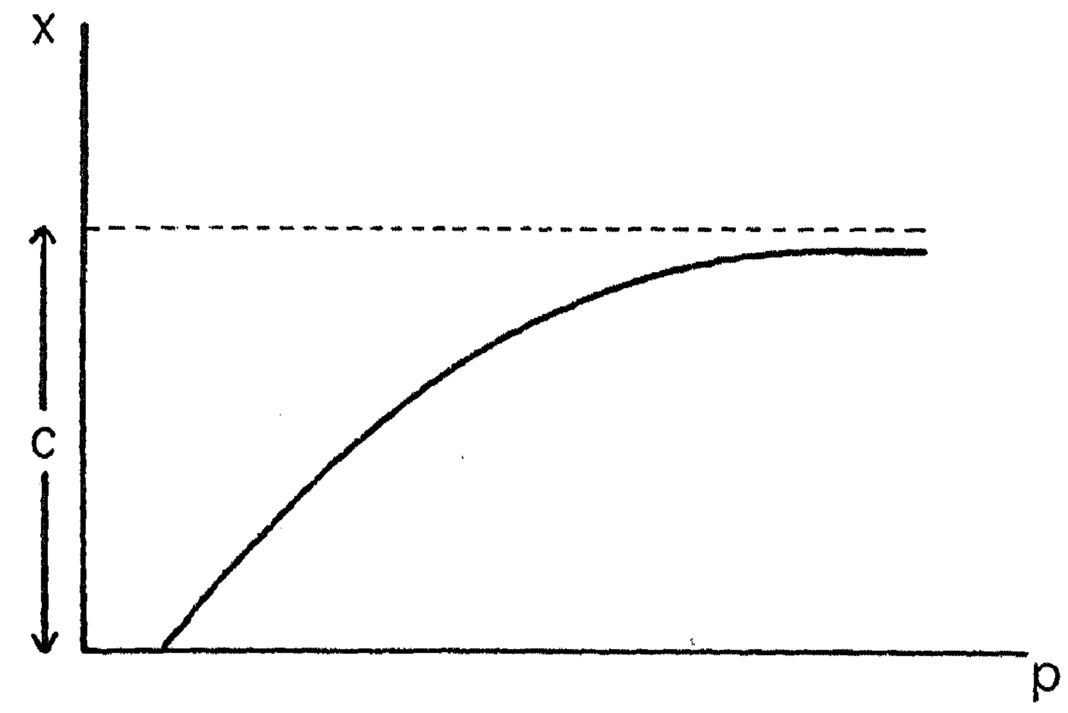

FIG. 6.—Example of a curvilinear relationship between two variables as represented by the supply curve (p = price, x = quantity supplied, and c = productive capacity).

commodity and the quantity supplied, x, on the assumption that there is a certain given productive capacity, c, and that no stocks are available. Starting from a relatively low price, every increase in price will initially produce a considerable increase in the quantity supplied. As the total capacity is approached, however, every further increase in the price by a constant amount will produce a smaller and smaller increase in the quantity supplied. The quantity supplied cannot exceed productive capacity and will approach this magnitude asymptotically; the horizontal line drawn at the distance c from the origin is called the "asymptote." In a time graph a curvilinear relationship of this nature will be recognized by the fact that the two series, though increasing and decreasing at the same time, show a pronounced difference in shape; the peaks in the p series will be much more pronounced and sharper than those in the x series. It will not be possible to make these series coincide by an appropriate adjustment of the scales. This phenomenon will, in particular, be encountered with respect to the prices and production of mineral raw materials in some periods of boom, when production bottlenecks occur.

We have so far discussed only the case in which the corresponding values of the two series referred to the same moment of time for the two economic magnitudes. If, however, there is a causal relationship between the two variables and if the process of causation which links Y to X takes a certain time, T, there would be good reason to compare the value of X at a given moment with the value of Y at a moment T later than the corresponding value of X; hence one would compare the value of X at time t with the value of Y at time $t + T$. If the two variables are plotted against time in the normal manner, the fluctuations of Y will be shown to be lagging behind those of X; by shifting the diagram of Y to the left over a distance T, the two series can be made to coincide again, provided that appropriate scales have been chosen for X and Y. Similarly, in the scatter diagram the co-ordinates should be chosen in such a way that the value of X at time t is combined with the value of Y at time $t + T$, in order to obtain points which lie on a straight line. In such instances we say that Y shows a lag with respect to X, or that the relationship between Y and X is a lagged relationship (Fig. 7).

In the case of two series with irregular movements, the presence of a lag can readily be observed empirically. If, however, the two series have monotonic movements, the lag is almost impossible to establish. If the movement is purely periodic, it is not possible to establish from the data the direction of the lag, that is, whether Y lags behind X or X lags behind Y, unless one has separate information concerning the order of magnitude of the lag.

Besides lagged relationships, many other relationships are possible between two variables in which time plays a role. A very common relationship in economic analysis is that of cumulation. Series Y represents the cumulation of series X, if the nth value for Y is equal to the sum of all values from 1 to n

for X. Hence, the first value for Y (Y_1) is equal to the first value for X (X_1); the second value, Y_2, equals $X_1 + X_2$; Y_3 equals $X_1 + X_2 + X_3$; Y_4 equals $X_1 + X_2 + X_3 + X_4$, etc. This relationship will occur, for instance, if Y represents the stock of a certain commodity, while X represents the excess of production over consumption for each time period. If X is gold production, then Y represents the total gold stock (assuming

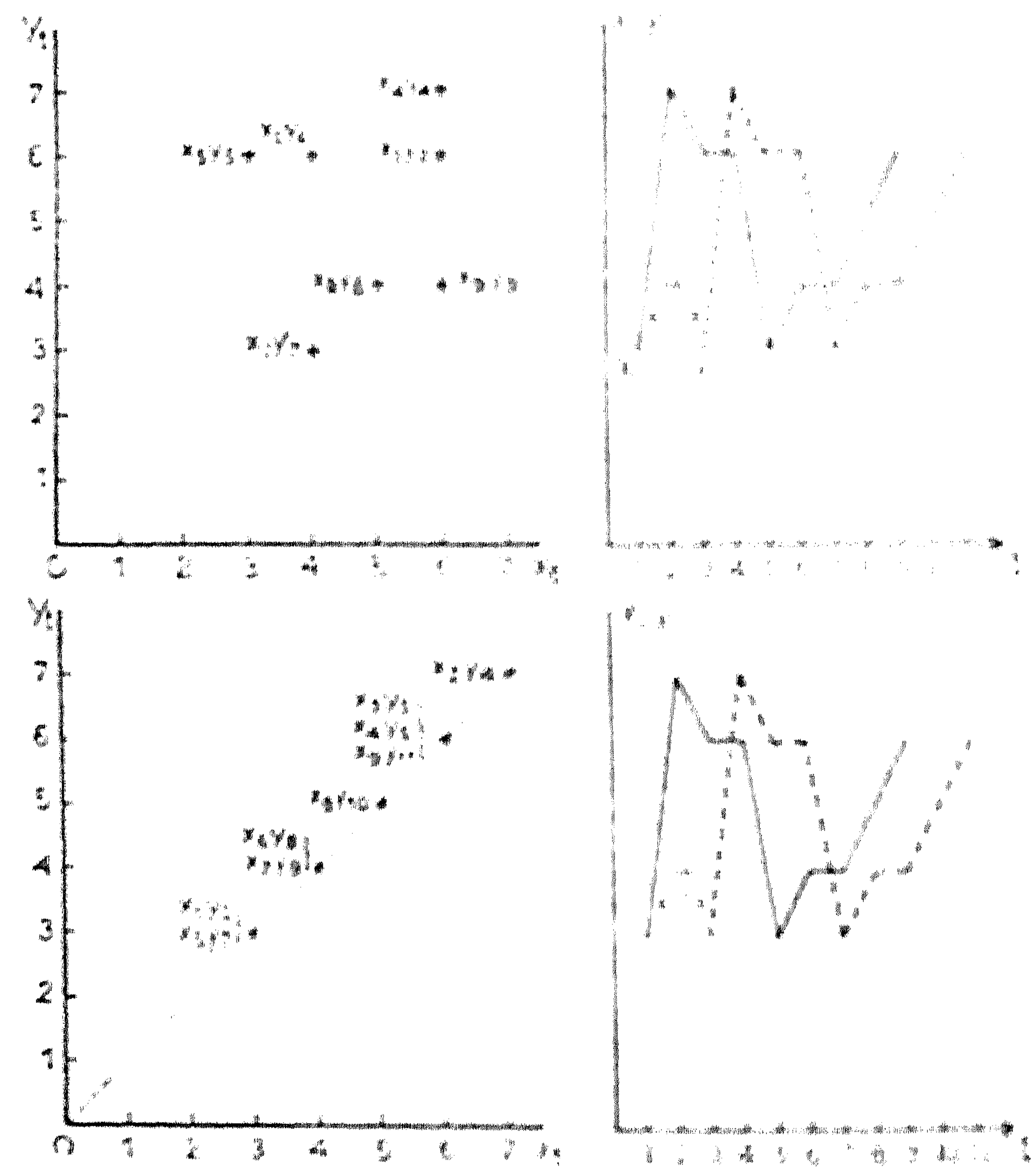

FIG. 7.—Example of a lag between two series X and Y. The top scatter diagram shows simultaneous value of X and Y; the bottom scatter diagram combines X_1 and Y_3, X_2 and Y_4, etc.

that the consumption of gold may be neglected); if X is the net increase of the number of houses during the year, then Y is the total stock of houses.

If X fluctuates, Y will fluctuate too. Normally, the fluctuations of Y will contain a trend component, unless the average value of X equals zero. If the series X is a sine curve with zero average, the Y curve will also be represented by a sine curve.

for X. Hence, the first value for Y (Y_1) is equal to the first value for X (X_1); the second value, Y_2, equals $X_1 + X_2$; Y_3 equals $X_1 + X_2 + X_3$; Y_4 equals $X_1 + X_2 + X_3 + X_4$, etc. This relationship will occur, for instance, if Y represents the stock of a certain commodity, while X represents the excess of production over consumption for each time period. If X is gold production, then Y represents the total gold stock (assuming

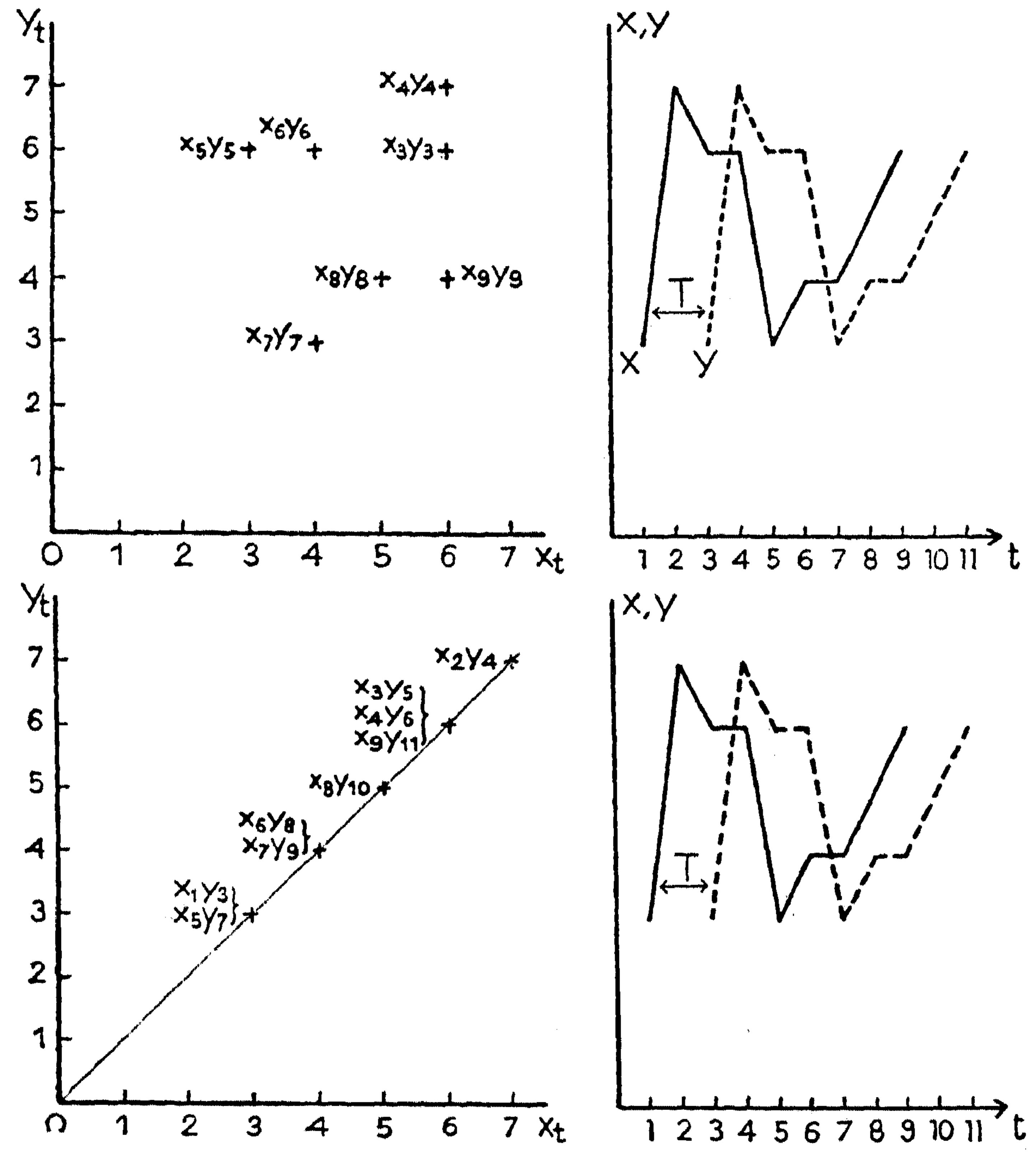

FIG. 7.—Example of a lag between two series X and Y. The top scatter diagram shows simultaneous value of X and Y; the bottom scatter diagram combines X_1 and Y_3, X_2 and Y_4, etc.

that the consumption of gold may be neglected); if X is the net increase of the number of houses during the year, then Y is the total stock of houses.

If X fluctuates, Y will fluctuate too. Normally, the fluctuations of Y will contain a trend component, unless the average value of X equals zero. If the series X is a sine curve with zero average, the Y curve will also be represented by a sine curve.

This may be proved mathematically and can also be seen from the following example computed on the basis of approximate values (compare Fig. 8):

t	= 0	1	2	3	4	5	6	7	8	
X	= 0,	0.7,	1,	0.7,	0,	−0.7,	−1,	−0.7,	0,	etc.
Y	= 0,	0.7,	1.7,	2.4,	2.4,	1.7,	0.7,	0,	0,	etc.

Although Y also shows a sine curve, it is lagged in comparison to X; the peaks of Y are approximately one-fourth of the period of fluctuation behind those of X. Any numerical experiment will show that the cumulation of an irregular series will not reproduce a series of the same shape.

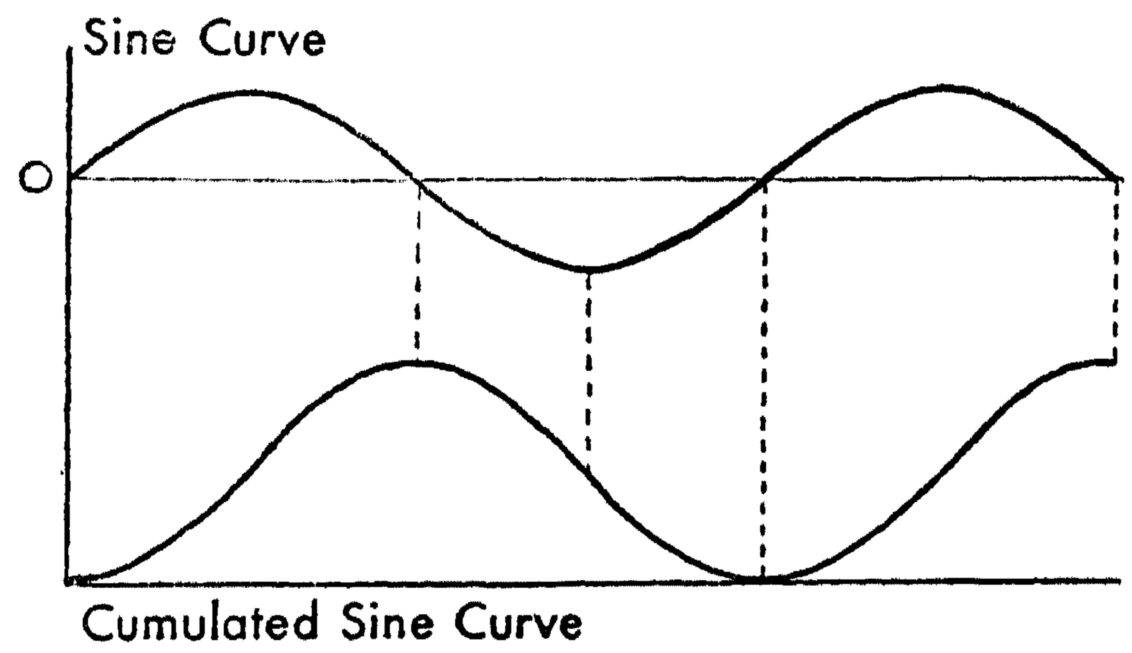

FIG. 8.—Sine curve and cumulation of sine curve

A purely empirical inspection of Figure 8 might lead one to believe that the relationship between X and Y was a lagged one, whereas in reality it is a cumulative relationship. Only theoretical knowledge of the real relationship between variables can prevent erroneous inductions of the type just mentioned.

THE RELATIONSHIP AMONG MOVEMENTS OF THREE OR MORE SERIES

In the preceding section we discussed the relationship between the movements of two series. The changes in two series will be exactly parallel when the changes in the one are caused by changes in the other and by nothing else. In the case of economic phenomena, however, fluctuations in one variable are usually due to fluctuations in more than one other variable. On this account it will normally not be possible to find a relationship between one of these causes and the variable in whose causation one is interested. Instead of the relationship between two variables, we will now have to have recourse to more complicated relationships. The simplest of these occurs when the values of one variable are equal to the sum of the simultaneous values of two other variables. The numerical example shown in Example I may make this clear; in each year the value of

EXAMPLE I

Year	1	2	3	4	5	6	7	8	9
Series A..........	0	+2	0	0	−2	−1	−1	+2	0
Series B..........	0	+2	0	+1	0	0	0	−3	0
Series C..........	0	0	0	−1	−2	−1	−1	+5	0

series A is equal to the sum of the values of series B and C. A relationship of this nature can graphically best be represented by time graphs. Figure 9 shows this for Example I. It will be noted that certain of the characteristics of series B and certain of those of series C can be found in series A. Thus, the first peak in year 2 can be traced back to series B and the second peak in year 8 to series C. The second peak is less pronounced in A than it is in C because it is in part offset by a trough in series B. The irregular depression in series A in years 5–7 is also found in series C. In short, series A shows the joint effects of series B and C. Neither series B nor series C can by itself explain series A; only their sum can explain it.

The relationship between series A which we want to explain and the explanatory series B and C can be much more complicated in many respects. To take a relatively simple case, it may be that series B and C require to be multiplied each by a certain coefficient before they are added up. An example of this is worked out below and is also shown in Figure 9, in which series A' equals $2\,B + \frac{1}{2}\,C$.

EXAMPLE II

Year	1	2	3	4	5	6	7	8	9
Series A'.........	0	+4	0	+1.5	−1	−0.5	−0.5	−3.5	0
$2\times$series B.......	0	+4	0	+2	0	0	0	−6	0
$\frac{1}{2}\times$series C.......	0	0	0	−0.5	−1	−0.5	−0.5	+2.5	0

It will be seen that the values of B are multiplied by 2 and those of C by $\frac{1}{2}$ in order to yield A' by addition. Graphically, the depression of A' in the years 5–7 can again be found in series C, and the peak in year 2 is now more pronounced in series B. A', however, has a peak in year 4 which A did not have and which is due to B; in the first example this peak was offset by a trough in C. In year 8 the greater influence of series B now dominates the peak in C; hence, there is a trough in A', whereas in our first example there was a peak in A.

It would be possible to cite many more complicated cases. In the examples given so far the relationships were additive, that

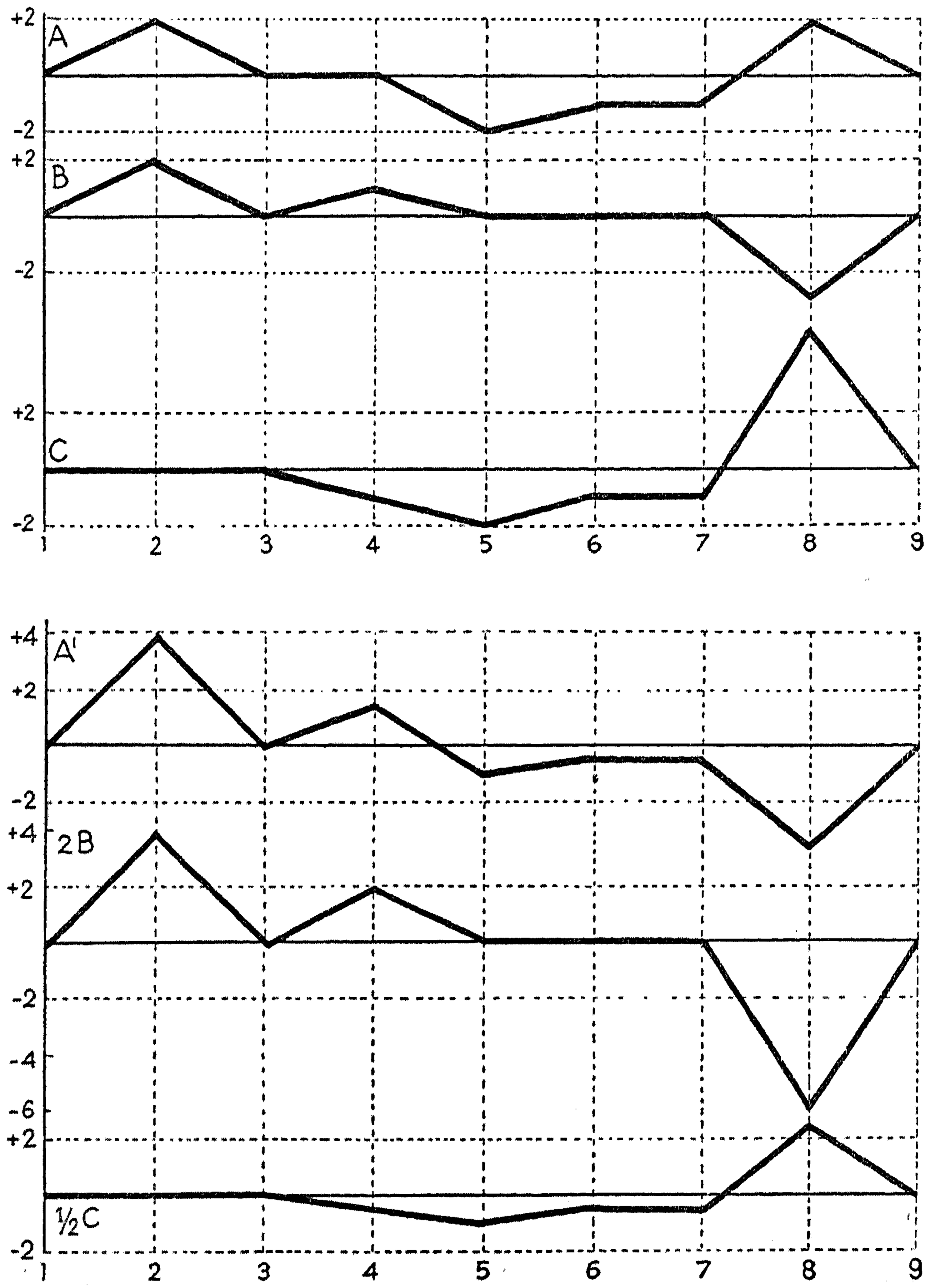

FIG. 9.—Examples of multiple correlation; $A = B + C$; $A' = 2\,B + \frac{1}{2}C$

is to say, A was obtained by the addition of B and C, A' by the addition of $2\,B$ and $\frac{1}{2}\,C$. Sometimes more complicated mathematical relationships (multiplicative, exponential, etc.) may prevail. It may, however, be proved that, for fluctuations that are relatively small in proportion to the average of the series, the more complicated operations will yield results only slightly different from the results of appropriate additions. Thus, for instance, the multiplication of two index numbers that deviate only slightly from 1, as shown in Table 1, may, with a high degree

TABLE 1

MULTIPLICATION OF TWO INDEX NUMBERS

Year	1	2	3	4	5	6	7	8	9
Series P..........	1.02	1.04	0.99	1.00	0.98	0.97	1.02	1.00	0.98
Series Q..........	1.01	0.99	.98	1.02	1.00	1.01	1.02	0.97	1.00
$P \times Q$...........	1.0302	1.0296	0.9702	1.0200	0.9800	0.9797	1.0404	0.9700	0.9800

TABLE 2

APPROXIMATION OF PRODUCT BY SUM

Year	1	2	3	4	5	6	7	8	9
Deviations:									
P.............	0.02	0.04	−0.01	0.00	−0.02	−0.03	0.02	0.00	−0.02
Q.............	.01	− .01	− .02	.02	.00	− .01	.02	− .03	.00
Sum...........	0.03	0.03	−0.03	0.02	−0.02	−0.02	0.04	−0.03	−0.02

of approximation, be replaced by adding to 1 the sum of the deviations of each of the two series P and Q, as shown by the figures in Table 2.

If the two series have other averages or if other mathematical operations have to be performed, the addition used as an approximation becomes somewhat different. But it will always be possible to give a suitable approximation by addition. For this reason, Example II has a very general significance. We shall therefore discuss it in slightly more detail.

The coefficients 2 and $\frac{1}{2}$ by which series B and C are multiplied may be called "influence coefficients." The series $2\,B$ and $\frac{1}{2}\,C$, given on the last two lines of Example II, are called the influence of B on A' and the influence of C on A'. If we desire to represent these influences by one single number, by which we may want to indicate whether the influence of B on A' is large or small, we can use for this number the average deviation of the standard deviation of $2\ B$ or $\frac{1}{2}\ C$, respectively, as is known from theoretical statistics. The theory of mathematical statistics provides an answer to the question of how to select the influence coefficients of the two series B and C in order to obtain as good an approximation as possible to any given series A''. This is of particular importance if we want to verify and give quantitative precision to a certain economic theory which would state that the movements of the variable A'' result from the movements of B and C but which does not indicate how large the relative influences are. This is the usual situation in economic theory, namely, that it can give qualitative but not quantitative indications about certain relationships. The coefficients found by the methods developed in theoretical statistics are usually called "regression coefficients." If A'' is exactly equal to the sum of a certain number times B and a certain number times C or, as we shall say, to the weighted sum of B and C (as is the case in Example I for series A and in Example II for series A'), then a perfect multiple correlation between A'', on the one hand, and B and C, on the other hand, is said to exist; if A'' is not exactly equal to the weighted sum of B and C for all periods of time, there is a certain degree of multiple correlation which may be expressed by the multiple-correlation coefficient. Reference is made to the textbooks on statistics for the conditions which have to be satisfied in order to estimate, with a stated degree of precision, regression coefficients on the basis of series of figures for A'', B, and C.

Example II, as shown in Figure 9, will be considered as the prototype of a multiple relationship among more than two variables. The number of explanatory variables need, however, not be limited to two; indeed, there may be any number. But, whatever their number, the explanatory series will always have to satisfy the condition that they must provide the explanation for the peaks and troughs in the series to be explained. Thus, not all the explanatory series can be straight lines. Applied to economic fluctuations, this would mean, for example, that it would not be possible to explain cyclical fluctuations on the basis of series which themselves show only a trend movement. This statement would not seem to be redundant since, for example, in certain theories the recovery phase of the business cycle has been attributed to an increase in the population. Yet this series shows almost exclusively a trend movement. Similarly, changes in labor productivity cannot provide an explanation for cyclical fluctuations, since for the economy as a whole labor productivity normally shows a straight line or only a very slightly curved line.

CHAPTER TEN

THE PROCESS OF LONG-RUN DEVELOPMENT

THE CONDITIONS OF FULL UTILIZATION OF RESOURCES

ACCORDING to economic statics, a given status of population, technical knowledge, and stock of capital goods, combined with given values for a multitude of technical, psychological, and other coefficients, will, under certain conditions that we will have to analyze further, determine the levels of production, employment, relative prices, incomes, etc. As we said earlier, the instrument of economic statics is, therefore, not a fully adequate, although an important, instrument to explain the trend of economic development. Economic development occurs to a considerable extent as a result of changes in population, changes in technique, and changes in the stock of capital goods. These changes lead to certain problems that fall outside the scope of economic statics and also outside the scope of comparative statics. What will be the curves that production, employment, incomes, etc., describe over time? What is the influence on them of the rate of increase in population, in capital formation, and in technical development? What effect do interruptions such as wars have on this development? Will the income distribution tend to change in such a way that it will lead to socially intolerable situations? We shall have to expand economic statics into a theory which encompasses the process of long-run development and to try to apply this theory to the solution of the problems indicated. Before we make any use of the theorems of economic statics, it may be useful to make clear certain of their assumptions which have not always been realized clearly.

Economic statics in its more usual versions is characterized by a certain optimism concerning the question whether all productive factors will be used in the process of production. Often the equilibrium level of production is portrayed as the result of the happy co-operation of all labor, all natural resources, and all capital goods. Very generally speaking, this proposition is made plausible in the following way. If any workers or capitalists are involuntarily excluded from participation in the process of production, they will offer their services at a lower wage or a lower rate of interest and will in this way insure their acceptance in the process of production. In other words, full employment of all productive factors is believed to be guaranteed by considering only the supply of the factors of production. However, it would appear from experience that the importance of the changes mentioned, that is to say, the inducement exercised on entrepreneurs by price-cutting of the factors of production, is overestimated. This is indicated by the periods of long-drawn-out depressions that have occurred particularly since 1920 but also in earlier periods. It would seem clear from experience that a certain "suction" on the entire economy is necessary to have it run at full capacity.

More concretely and more exactly: the symmetry which is assumed in most conventional versions of economic statics between the demand and supply sides of the market would not appear to be in correspondence with reality. In particular, the volume of production of goods and services is determined mainly by the demand for them. In most markets the initiative concerning the quantities is with the buyers, that concerning the prices with the sellers.

We may go somewhat more deeply into this question and analyze at whose initiative important changes in the volume of total production are brought about. Total production may for this purpose be considered to consist of the production of consumption goods and the production of investment goods. For convenience we count services among goods. The demand for consumption goods is exercised to a very large extent by persons receiving moderate incomes and whose demand is determined by their income. They cannot spend much more, and they do not want to spend much less, than their income. The magnitude of their income is almost exclusively determined by the volume of total production and does not, therefore, add a new element to the determining factors. The consumption expenditure of persons receiving higher incomes has a greater degree of freedom; but it, too, is affected by the size of these incomes. Most independent, therefore, are the expenditures for investment. Except for the investment expenditures of small entrepreneurs who have no access to the capital market, investment expenditures are not limited by the size of incomes but can exceed them by recourse to credit.

The existence of strong motives to make investment expenditure is, therefore, of great importance in the determination of the total volume of production. If we look at the nature of large segments of investment expenditure, we are led to believe that a considerable influence is exercised by new technical possibilities, for instance, in railroad investment, electric-power stations, and the combustion engine, and by the discovery of natural resources in new countries. The nineteenth century has been rich in such new possibilities, which may have been responsible for keeping total demand for commodities at a relatively high level, thus providing the necessary "suction" to which we referred and the constant tendency to the use of all, or at least a very great part, of the productive resources of the economy. It may also be, as some have maintained, that the general mentality of entrepreneurs is subject to changes and that periods of greater optimism and audacity alternate with periods of caution and timidity. These differences have been associated with differences between succeeding generations and with the alternating domination of the male and female elements. Such suggestions, though interesting, do not appear to be very strongly founded scientifically; in any case, it is not the task of economics to analyze them further. These are data for economic science, and we cannot probe more deeply into them.

In any case, it will be necessary in our analysis to make a sharp distinction between two cases, that in which the total demand for commodities is intensive and that in which it is weak. In a period of intensive demand one may say approximately that the volume of production is determined by productive capacity; in periods of weak demand the influence of productive capacity is minor.

Both situations may be considered as special cases of one and the same demand and supply diagram in which the level of the demand curve changes (Fig. 42). According to usual practice, the quantities of production are indicated along the horizontal axis, prices along the vertical axis. The solid line S indicates the supply curve; the distance c indicates productive capacity. As the price increases from a very low level, only a very few enterprises will initially be able to produce; supply will approximate

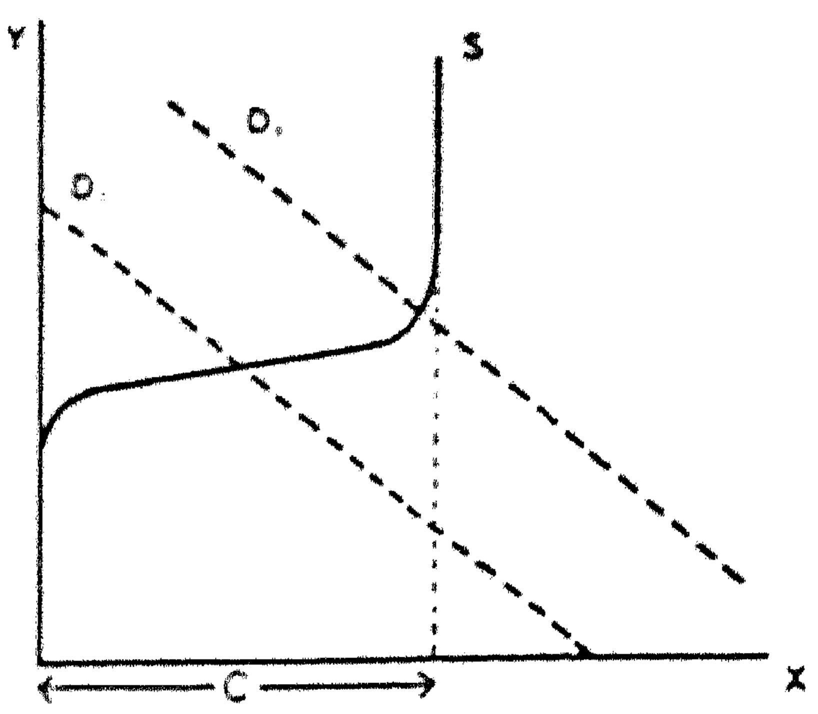

FIG. 42.—Influence of the intensity of demand on the quantity demanded and the price. In the case of strong demand (D_1), the quantity sold equals c (capacity); in the case of weak demand (D_2), c has no effect.

capacity only when the price has risen above the cost of production of the majority of producers. Obviously, supply cannot exceed capacity. The dotted lines, D_1 and D_2, indicate two positions of the demand curve. In case 1 there is a strong, in case 2 a weak, demand. In case 1 the quantity produced and sold is practically equal to c, whereas in case 2 this is not so; in case 2 the size of c is of minor consequence; but the level of the cost of production of the majority of enterprises is of importance in the determination of the quantity produced and sold.

SIMPLIFYING ASSUMPTIONS

We shall now analyze in more detail the contents of economic statics on the assumption of a strong demand. To repeat, our purpose in this analysis of statics is to gain an insight into the process of development and into the problems connected with it. In this context we shall apply simplification to all aspects that are not relevant to our purpose; if we did not, the system would become unmanageable. For instance, we shall not treat the markets for individual commodities separately but deal with only one product, which is assumed to be suitable for both consumption and investment purposes. We shall also distinguish only one type of labor and one type of capital goods. Often, we shall include the productive factor, nature, among the capital goods; in practice the distinction between nature and capital goods is in any case not an easy one. We shall at most consider only one type of land. We shall disregard the distinctions between the successive stages in the process of production, as well as the lack of mobility of labor and capital. We shall also leave out of consideration the difference in profitability of different enterprises. Some of the distinctions which fall victim to our simplifications are of importance primarily for the explanation of short-term movements; these movements are not the object of our present analysis. In the same train of thought, we shall leave out of consideration all phenomena of a speculative nature, such as the response of entrepreneurs or consumers to the rate of change in prices, as distinguished from the level of prices.

These simplifications raise a question of methodology on which we can touch only lightly. Is it permissible to disregard the phenomena mentioned? Is not the long-run development to some extent determined by them? A closely related question is whether any meaning can be attributed to the concept of a static equilibrium in a society that is subject to cyclical fluctuations. The following would appear to be a rough, provisional answer to these two questions. If the short-run fluctuations, in particular cyclical fluctuations, are damped or undamped but not antidamped, the long-run development is not affected by them, and in our study of the static equilibrium we may in this case legitimately disregard phenomena that are of importance only for the short-run fluctuations.

To offset these simplifications, we may enrich our analysis in directions that are of particular importance for our present purpose. Thus, for instance, we will not now consider the stock of capital goods as given but as a result of a process of accumulation: it is the sum of net investment over the entire period of the preceding years. We shall also admit changes in the data over time, according to certain laws. This will apply in particular to changes in population and in technical knowledge.

With respect to a number of other points, we shall work according to the well-known method of decreasing abstraction. This will apply in particular to the functions of money in the economy. We shall start out by considering an economy without money. We shall then introduce money first as a unit of account and then as a means of payment. Finally, we shall introduce money as an instrument of saving. There is much to be said for an analysis that starts out from an economy without money. Many disturbances of economic equilibrium that result from the existence of money, such as the "money illusion," are of short-run importance only; after a certain period the economic subjects will adapt themselves to changes in the value of money. If one considers only long-run developments, no great mistake is made by assuming that the adaptation to the changes in the value of money is immediate.

Within our analysis of an economy without money, we shall start again with the simplest case, following Cassel, in which we consider only increases in population and "horizontal" increases in the stock of capital goods; we shall subsequently take up the increase in capital intensity and only finally changes in technical knowledge.

THE BASIC RELATIONS OF THE SIMPLIFIED MODEL

In the economy simplified in the way indicated, two markets must be distinguished, namely, the labor market and the capital market. In the labor market there is an exchange of labor against commodities. In the capital market there is an exchange of claims to capital goods against commodities. Both markets are therefore also markets of commodities. The economic variables which relate to these markets and the development of which should be explained by economic theory are (*a*) the prices, namely, the real wage rate and the real rate of interest, and (*b*) the quantities traded, that is to say, the quantity of labor used, the quantity of capital used, and the volume of production. The quantity of labor used may also be regarded as the level of employment. The real rate of interest should be considered as the income of the owner of a unit of capital goods, expressed in terms of consumption goods. For simplicity, we shall drop the adjective "real" when we speak of prices. The level of prices and the magnitude of employment and of the use of capital are determined by the demand and supply functions in the two markets. The volume of production depends, further, by a technical relation, on the quantities of labor and capital used. This latter relation we will call the "production function."

a) The supply of labor

The supply function of labor indicates the number of workers who will offer their services at various wage rates. The most important supply factors affecting this function are the size of the population and the level of real wages considered normal or necessary. The larger the population, the larger will be the supply of labor, all other things being equal. The higher the conventional standard of living in a certain country, the lower will be the appreciation of a certain given real wage and the volume of labor supplied at that wage. The length of the working day may be considered as a third supply factor. One may want to go into further detail and consider, for instance, in addition to the size of the population, its composition, such as the distribution over productive and nonproductive age classes. These further details, however, need not be incorporated in our analysis.

A crucial question with respect to the supply function of labor is its slope. It is theoretically conceivable both that a higher real wage rate will increase the supply of labor and that it will decrease it. A higher wage will make employment more attractive and might therefore lead to a higher supply of labor; for the same reason, a lower wage might lead to a lower supply. This relation will be quite pronounced, for instance, if labor is unionized and would strike in case of a reduction of wages. It is also possible, however, that a higher wage will induce fewer members of a family to seek work because a certain minimum income, considered necessary, can be obtained with a smaller amount of work. For some members of the family the value of leisure may be greater at a higher wage than the wage which could be earned. This phenomenon is observed particularly in the case of colonial workers; in times of prosperity, however, it also applies to large families in Western economies.

Since both possibilities can exist theoretically, only observation can decide which of the two applies in a particular case. Very little empirical work has been done in this field, and the results do not always point in the same directions. In a number of cases Professor Douglas has found that an increase in wages was accompanied by a reduction in the supply of labor; this applies in particular to women and younger persons. Other studies, however, have shown opposite results. In theoretical studies of a more general character, concerning the long-run process of development, it has often been assumed that the influence of the wage rate on the supply of labor may be disregarded; it is assumed, then, that the entire population that is fit to work will desire to work.

b) The supply of capital

The supply function for capital indicates what amount of claims to capital goods will be supplied at any given rate of interest. This supply function has a higher degree of abstraction than that of the labor market. Suppliers in this market do not own the capital goods themselves but only claims to them. In reality, a considerable part of the supply of claims to capital goods is not under the control of the owners of these claims; the disposition over these claims is often made by entrepreneurs rather than by the owners. Only when enterprises attract new capital do the owners of claims decide whether they want to make the additional capital available to production. In line with our simplification we assume that the owners of capital goods receive a remuneration of interest in terms of commodities.

The supply will depend, in the first place, on the total stock of capital goods available. Here, as in the case of labor, it is uncertain in which way an increase of the remuneration, that is to say, of the rate of interest, will affect supply. Empirical studies are necessary in this case, too, to determine concretely in which direction the effect is. Little positive information is available at present; as in the case of labor, it is not unusual to assume that changes in the rate of interest will have only a small effect on the supply of capital. We shall make that assumption, and we shall consequently assume that, in times of intensive demand for commodities, the entire available stock of capital goods will be offered. Only very recently have tendencies toward monopoly on the part of savers made themselves felt; their influence has on the whole been so small that we may disregard it.

c) The production function

The demand for labor and for capital is exercised by the entrepreneurs. We may assume that the entrepreneurs are motivated by a desire to maximize their profits by the organization of production. They combine certain quantities of labor and of capital, in order to produce certain quantities of products. The yield of production, in real terms, will have to be used to pay wages and interest, both also in real terms; the remainder is the profit of the entrepreneurs. In order to determine the demand function of the entrepreneur, we have to ask this question: Up to which point will the entrepreneur continue to demand labor and capital, given a certain wage per unit of labor and a certain rate of interest for each unit of capital? He will clearly continue to purchase labor until the yield, in terms of product of additional labor, is no longer in excess of the wage, in other words, until the last worker no longer makes an addition to profits. The yield of the last worker is identical with the marginal productivity of the total quantity of labor employed. The marginal productivity will thus be equal to the wage rate. Similar considerations will lead to the conclusion that the quantity of capital employed will be such that the marginal productivity of capital is equal to the rate of interest.

The marginal productivities of labor and capital introduced in the preceding paragraph follow from the production function. This function indicates the quantity of product obtained by the application of given quantities of labor and capital.[1]

1. Followers of Boehm-Bawerk may prefer to state that the quantity of product depends on the quantity of labor and the degree of "roundaboutness" of production. The latter, however, may be derived from the quantity of capital; hence, there is no contradiction between the two ways of exposition.

Let us analyze the production function in somewhat greater detail. The definition just given has to be qualified in the sense that the application of labor and capital should be optimal. This implies that production takes place in enterprises of optimal size. We assume that the economy we consider is so large that the optimal enterprise is small in comparison to the total economy. If this condition is satisfied, it will be possible to make a small addition to total production by the addition of one new enterprise. In this situation we have to be concerned only with production functions in which the factors of production are combined in the optimal way; the extent of production that is taking place below the optimum may be disregarded because the total quantity of product which is produced in these circumstances—as the last enterprise does not work at full capacity—can only be small in comparison to total production. If these conditions are satisfied, a proportional increase of the quantities of labor, capital, and land employed (for instance, all increased by 10 per cent) will lead to an equiproportional increase in the quantity of product.

The production function will be quite different for different industries, and it may also be different for different enterprises within the same industry. It is also reasonable to assume that it changes over time. For our purposes, however, we need a production function of a rather general character, that is to say, one which would be applicable to the total production in an economy. Such a production function is of little interest to individual enterprises, and it is perhaps for this reason that relatively little research has been done with respect to it. Only a few research workers, in particular Professor Douglas,[2] have tried to establish production functions of this character on the basis of statistical data.

The simplest formula one might imagine is one in which the elasticities are constant. This is the formula which Professor Douglas has assumed for industry as a whole. The elasticity of production with respect to a certain factor of production, labor or capital, is defined similarly to the definition of elasticity of a demand or supply function, viz., as the ratio between a relative

2. Paul H. Douglas, *The Theory of Wages* (New York: Macmillan Co., 1934).

increase of the volume of output to a small relative increase of the factor of production which causes it. If an increase of 1 per cent in the quantity of labor were to produce an increase of 0.7 per cent in the quantity of product, the elasticity of production with respect to labor would be indicated as 0.7. Even if a constant elasticity function were not strictly applicable, it could still be used as an approximation for relatively small changes in the quantities. The changes that have taken place during the past century have, however, by no means been small. But the studies by Douglas do not give any indication to the effect that the elasticity would have changed considerably. If all enterprises are of optimal size, the sum of the elasticities of production with respect to the different factors of production should be equal to unity. If the three factors of production all increase by 1 per cent, the total volume of product will, as we have seen, increase also by 1 per cent, under the assumption made. The increase in the quantity of product is equal to the sum of the elasticities; thus, these elasticities together should be equal to unity. The following estimates for these elasticities have been found: for labor approximately 0.7, for capital approximately 0.2, and for land ("nature") approximately 0.1. This would imply that an increase in the application of the quantity of labor by 1 per cent, keeping the quantities of capital and land unchanged, would lead to an increase in the volume of output of 0.7 per cent, etc.

It would follow from this proposition that the percentage increase in the volume of production is equal to the weighted average of the percentage increases in the factors of production, with the elasticity coefficients as the weights.

The assumption of a constant elasticity, independent of the quantity of labor applied, would imply, strictly speaking, that there is complete substitutability among the factors of production, so that it would be possible to reduce the quantity used of any factor to as small a positive value as one likes. But it will never be necessary to carry this assumption to such an extreme point in the explanation of actual events.

At the other extreme, one might assume that the factors of production were strictly complementary. This would imply that for any given quantity of product the quantities of the factors of production required are entirely fixed and that it is not possible to change them, for instance in response to changes in the relative prices of the factors of production. Even with full complementarism, changes would, of course, be possible as technical knowledge changed. In the short run, a situation of strict complementarism may be approximated; the longer the period available for adaptation, the greater will be the substitutability.

From the assumption of constant elasticities of the volume of production with respect to the different factors of production, it would follow that the total national product will always be distributed in the same proportions over the various factors of production. An increase in the number of workers would, under that assumption, not lead to a greater share of workers in the national product; their larger number would be compensated by a lower wage rate, and in the case of a constant elasticity the two changes would exactly offset each other. The fact that the distribution of national income over wages, interest, and rent varies relatively little among countries and among periods would seem to indicate that the assumption of a production function with constant elasticities constitutes a reasonable approximation to reality. The actual figures of relative shares depend on the definitions selected, in particular on whether one includes in "labor" only manual or also intellectual and organizing labor. If labor is defined in the widest possible sense of the word, it is usually found that approximately 70 per cent of the national income is labor income, approximately 20–25 per cent capital income, and 5–10 per cent rent income.[3]

The demand and supply schedules for labor and capital sketched in the preceding paragraphs determine the prices and quantities used of these factors of production, provided that all the data in these schedules are given. The production function will then determine the quantity of commodities produced, which in our simplified economy would be equal to the real national income. The data of the economy would include the population, that wage rate which is considered normal, the length of the working day, the stock of capital goods available, the differ-

3. Cf. Table 8, p. 37, *supra*.

ent elasticity coefficients, etc. In the special case in which the elasticities of supply of labor and capital are equal to zero, the volume of employment would be equal to the size of the working population, and the quantity of capital used equal to the total stock of capital; these, with the production function, would determine the volume of production. The wage rate and the rate of interest would be equal to the marginal productivity for the given quantities of capital and labor. This would determine all economic variables in our model.

d) The formation of capital

We have stated already, however, that we will not consider the stock of capital goods as given but rather, in contrast to economic statics, will consider its explanation as one of our objectives. The total stock of capital goods may be considered as the sum of the net additions to the stock of capital goods in successive years, that is to say, as the cumulation of net real investment. We must, therefore, find out what determines the latter. It results mainly from savings and is therefore dependent on the use made of income. Savings will, first of all, depend on the magnitude of national income; the higher the income, the larger will be the amount saved and the greater the addition of capital goods to the existing stock. The fraction of income saved has remained relatively stable over time; as a first approximation it may therefore be stated that investment is proportional to income. On closer analysis, however, the propensity to save is found not to be a constant. It results, for the economy as a whole, from the propensities to save of individual persons, and these propensities depend to a large extent on the levels of individual incomes. As the average of individual incomes rises, the propensity to save will also rise. It may also depend on the rate of interest but, as we noticed, this influence is not great. There may be a number of other factors, all of which, however, are of secondary importance.

Investments may also be financed initially by the creation of credit; in real terms applicable to our model this implies that in addition to voluntary saving there is also a certain involuntary saving. Forced savings occur particularly in periods when there is a strong tendency to invest. In times of a weak tendency to invest, not all savings are absorbed; but we are not dealing here with these latter periods. In periods of a strong tendency to invest there will be a greater increment of the stock of capital than is compatible with the normal rate of saving. But there is no indication that there have been any great fluctuations in the tendency to invest during the course of the nineteenth century and the first decades of the twentieth century, excepting perhaps a certain relaxation in the period from 1919 to 1939.

Savings are made primarily from incomes received from capital since in the main these are the largest incomes. If we may assume that the production function has constant elasticities, income from capital will be a constant proportion of total income.

As a first approximation, it is therefore reasonable to assume that capital formation is proportional to income, that is, to production. As a closer approximation, it may be stated that the increase in the stock of capital will be slightly more than proportional to the increase in the volume of production.

ANALYSIS OF ECONOMIC DEVELOPMENT

a) Horizontal growth

Using the instruments developed in the preceding section, we now endeavor to explain the process of economic development. We may start with a simple case, already treated by Cassel.[4] We shall call this case "horizontal growth." This term refers to a case in which there is no change in the ratios between the quantities of labor, capital, and land used; the formation of capital and the exploitation of new natural resources are used only to provide for an increase in population at a constant level of consumption. Horizontal growth will occur if, at a certain period and perhaps by accident, the rate of saving is such that the relative increase in the stock of capital is equal to the relative increase in the population. If the population increases by a constant percentage per year, the population curve will be an exponential curve. If the population follows a growth curve, that

4. Gustav Cassel, *The Theory of Social Economy* (New York: Harcourt Brace, 1932), Book I, chap. vi, "The Uniformly Progressing State."

is to say, a curve with a declining rate of increase, and capital formation runs parallel, total production will also show a declining rate of increase.

With horizontal growth, prices will remain unchanged. At any one time, the economy will be an exact replica of the economy at an earlier time, with all quantities multiplied by a certain factor.

b) Capital formation

As our next slightly less simple case we consider an economy in which the rate of capital accumulation is more rapid than that of the increase in population. For convenience we assume that the quantity of land increases in proportion with the population; we assume, further, that there is no change in technical knowledge. In particular, the production function remains unchanged. However, there will now be a change in the relative quantities of labor and capital used; production will become more capital-intensive. Mechanization, the replacement of labor by capital, will occur. Since the increase in the volume of output will be equal to a weighted average of the rates of increase of the factors of production with the elasticities as weights, total production will now also increase more rapidly than the population. Per capita income will increase. As we have seen, capital formation may be considered as approximately proportional to the volume of production. It will increase, therefore, at the same rate as production, in other words, at a slower rate than the stock of capital goods. The latter, therefore, will increase by decreasing percentage amounts. Production will no longer follow an exponential curve. The exact nature of the curve which it will follow cannot be discussed without introducing certain mathematical complications.

At the next stage we assume that the state of technical knowledge increases. A technical improvement may be defined as a change in the function of production in such a way that the same quantities of labor, capital, and land will produce a greater volume of output. This may also be expressed in this way: the same quantity of product can be obtained with a smaller quantity of the factors of production. The former expression, however, is the simpler, particularly for the case to which we want to apply it. Technical improvement may occur both if substitution between the factors—labor, capital, and land—is possible and if the factors are fully complementary. In the former case the second description has to be used. A smaller quantity of the factors of production has to be interpreted, then, as quantities which together cost less than the quantities used previously. It may well be that larger quantities of one or two of the factors are used, provided that the higher costs are more than compensated by lower costs for the other factor or factors.

We shall concentrate, for the moment, on the case in which the factors of production are always fully employed and in which, moreover, some substitution is possible among them. In that case, technical improvement will always lead to a larger volume of output. The volume of production will increase more rapidly if we assume a continuous improvement in technical knowledge than without such improvement. Assuming, again, a capital formation proportional to the volume of output, the former will also increase at a more rapid rate. It is now again possible that the percentage increase of income and of capital formation is the same as that of the stock of capital goods, so that the volume of output may again follow an exponential curve. In that case, output will increase more rapidly than the population, if the latter follows a growth curve. It would seem that this case provides a good approximation to reality.

It is clear from these simple examples that production may follow different curves, among them exponential curves but also very many other curves. It is also clear that the rate of increase of population, the rate of capital formation, and the rate of the development of technical knowledge influence the shape of this curve. These influences may be calculated numerically in certain simple cases, but in more complicated cases, particularly if account has to be taken also of changes in wage demands and of an elasticity of the supply of labor different from zero, the problem can be solved only by mathematical treatment.

In the last case it will follow from the production function that the percentage rate of increase of production per annum would be equal to the sum of (*a*) 0.7 times the percentage increase of the population, (*b*) 0.2–0.25 times the percentage increase of the stock of capital goods, and (*c*) 1 times the percentage increase of production as a result of technical improvement, that is to say, the increase of production which occurs if the quantities of capital and labor remain unchanged.

The percentage increase under (*b*) may be related further to the propensity to save, which is a more basic datum than the percentage increase in the stock of capital goods. If the ratio of the value of the stock of capital goods to the annual national income is assumed to be 5—a rather realistic figure—then we may substitute for (*b*) 0.04–0.05 times the propensity to save (expressed as the percentage of national income saved). In order to obtain an impression of the relative importance of the three components, one should know the percentage increase in the population, the propensity to save, and the increase in production as a consequence of technical improvements. Certain provisional measurements (of which the latter magnitude is the most provisional) would lead to the following numerical values. In the period of four decades preceding the first World War the rate of increase in the population in the four largest industrial countries was on the average about 1 per cent. The propensity to save may be put at 10–15 per cent of income; the increase of production as a consequence of technical improvement amounted to about 1 per cent per year. On the basis of these figures the three components indicated would be as follows: (*a*) 0.7 per cent to be attributed to the increase in the population, (*b*) 0.4–0.8 per cent to be attributed to capital accumulation, and (*c*) 1 per cent to be attributed to technical improvement.

It is clear from the preceding reflections that the relative significance of these components may be changed very considerably, for instance, if the rate of increase of the population slows down; the coefficient would also be affected if the elasticity of supply of labor were different from zero.

If the elasticity of supply of labor were equal to -1, the propensity to save would have no influence whatsoever, whereas the influence of the population increase would be greater; if, on the other hand, the elasticity were equal to 0.5, the influence of the propensity to save and of technical improvement would be twice as large and that of the rate of increase in the population would be less than the one given above.[5]

DISTRIBUTION OF INCOME

After dealing with the development of production, we may now turn to the wage rate and the rate of interest. As we have seen, they will be equal to the marginal productivity of labor and capital, respectively. In the case of horizontal growth, neither of these will change, and the wage rate and the rate of interest in real terms will remain constant. If the capital intensity increases without technical improvement, the marginal productivity of labor will increase and that of capital will fall. Technical improvement, however, will add an increasing component to these two movements. Technical improvement would also make possible an increase in the rate of interest. In actual fact, the wage rate has shown a sharply increasing tendency, whereas the rate of interest seems to be falling in the long run.

We may now come to the question of the changes in the distribution of income. We have seen that under free competition, with constant elasticities of the production function, the distribution of income over the different factors of production would be in a constant proportion. This proposition does not depend on the existence of free competition in the labor and the capital markets, since it follows directly from the proposition that entrepreneurs will employ labor only when and up to the point where the increase in output is in excess of the wage. Income per head of the labor population, in comparison to income per head of other groups of the population, will depend on the number of persons in each group. If the number of persons in the working class increased more rapidly than that of the owners of capital and land, then the wage rate per head would develop relatively unfavorably. In practice, this question is complicated by the fact that one person may receive both labor and capital income. Our proposition, however, does indicate in any case that income per head is determined not only by economic factors but also

5. For a systematic treatment on a mathematical basis, reference is made to J. Tinbergen, "Zur Theorie der langfristigen Entwicklung," *Weltwirtschaftliches Archiv*, LV (1942), 511.

by population factors. The tendency shown by the labor population in the last decades toward a smaller rate of increase would tend to be beneficial to their per capita income, assuming no unemployment.

If it is correct to assume that the elasticities of the production function are approximately constant, it will follow that attempts to raise the wage rate by association on the part of the workers could not increase the total labor income. Such attempts would, on the other hand, increase total labor income if labor and capital were complementary factors of production. In that case the demand for labor would be determined by the total volume of capital and would, within certain limits, be independent of the wage rate. This applies to some extent to changes in the short run. In these circumstances an increase of labor income would be possible at the expense of the incomes of capital and land. In the longer run, however, account would then have to be taken of the following facts: the lower incomes of capitalists would lead to a lower rate of capital formation, resulting in a less favorable development of the demand for labor at a later time. It might be that, in the longer run, total labor income would be lower than it would have been without the increase in the wage rate.

If the production function had rigidly constant elasticities, an increase in the share of labor in national income could be achieved only by taxation policy and by various social policies.

It is to be noted that all this refers to the distribution of incomes and not to the absolute level of incomes. The latter might be considered the more important; the distribution of income does, however, also have a considerable importance, in particular because it is mainly responsible for tensions between different classes of the population. An increased total income, accompanied by an increase in the inequality of distribution, may have dangerous consequences for the stability of society.

It has to be borne in mind that our knowledge of the production function is limited and that there are no clear facts to prove that the elasticities of this function are in fact constant. If they are not, any conclusions based on their constancy will, of course, have to be revised.

CONSEQUENCES OF WARS

Following approximately the chronological order in which some of the basic problems have presented themselves to our Western society, we will now, after our treatment of the social problem, devote a few paragraphs to the influence of a large war on the long-run process of development. We do not deal here with the problems of the war economy as such (they are treated in chap. xi) but rather with the consequences of wars on the long-run process of development.

Seen from that point of view, wars have the following characteristics. Human lives and capital goods are destroyed; as a consequence, smaller quantities of the production factors, labor and capital, are available. In addition, the regular accumulation of capital goods is interrupted for a certin period of time. In two respects, therefore, the stock of capital falls behind what it would have been without a war. The remaining population is thus likely to have a smaller stock of capital goods per head. The money side of the process of accumulation has continued, however; the population has continued to make savings, in terms of money, which have been invested in government obligations for the financing of the war. At the end of the war these savings are not matched by capital goods; there is, nevertheless, the obligation on the part of the community to make available to the holders of these government obligations a portion of the national product. We may say that there is a fictitious capital which participates in the distribution of national income but not in its production.

If the supply of labor and capital were entirely inelastic (our simplest hypothesis), all labor and all real capital would participate in production and would yield a volume of output per head lower than before the war. Of this output, part has to be ceded to the owners of fictitious capital, leaving even less for the productive factors.

If the supply functions are not entirely inelastic, the situation will be more complicated. If the elasticity of supply of labor is negative, more work will be done because the heavy burden on account of the service of fictitious capital will reduce the worker's wage. If the elasticity is positive, there will be the opposite result: since less is paid for work, it will be less attractive. The same may apply to entrepreneurs who, in view of their reduced income, especially marginal income due to high and progressive taxation, may relax their efforts.

Much concrete research will be necessary before all the consequences of the accumulation of fictitious capital have been established; it would appear clear, however, that the existence of large amounts of fictitious capital created by the succession of two large wars may entail great dangers with respect to the volume of output.

STAGNATION

So far we have analyzed the process of development on the assumption of a continuous intensive demand, particularly for investment goods. On this assumption the volume of output was primarily determined by the capacity of production.

An entirely different picture is obtained when the autonomous demand factors are not strong. In that case, the capacity of production will be no longer the determinant of the volume of production; part of the capacity may remain unused. This situation may be described in the following simplified way: the total volume of production is determined by demand, and the total use of the factors of production is determined by the attempt on the part of the entrepreneurs to meet this demand in the cheapest possible way. Given the volume of production, the choice of the quantity of one or the two factors of production is free; the quantity of the other factor of production is then determined by the production function, in accordance with the principle of the lowest total cost. It is then possible that neither the total quantity of labor nor the total quantity of capital will be fully absorbed. Under this assumption the rate of development is relatively independent of the level of investment. The rate of increase in the population has also only an indirect and not very important influence on the development of production. The volume of production will now develop primarily as a result of changes in the autonomous demand factors. In the years since 1918 autonomous factors in a number of countries have tended to produce a low level of general demand. It appeared as if the opening-up of new territories were either impossible or politically too risky, as if all new inventions were in the direction of economizing capital, and as if a general tendency toward economy tended to stifle all public expenditure.

These factors must be considered as responsible for the high average level of unemployment in some of the Western countries in the interwar period. They were, however, not the only causes. Shifts from these countries to the younger industrial countries on the periphery of the world economy were also in part responsible. We do not want to discuss the actual details of this development within the limits of this book.

Let us see in somewhat closer detail in which ways unemployment can occur, in situations both of intensive and of weak demand. We shall make a distinction between voluntary and involuntary unemployment. Voluntary unemployment occurs if part of the working population does not offer its services for production. This will occur when the supply of labor has a certain elasticity. In the case of a negative elasticity, voluntary unemployment will occur when the wage rate is relatively high; leisure, in particular that of women and children, is preferred to additional income. In the case of a positive elasticity, voluntary unemployment will occur when the wage rate is relatively low; the wage offered is not considered sufficiently attractive to compensate for the additional disutility of labor. This latter situation may occur in particular when the population has other sources of income at its disposal, e.g., unemployment assistance. It may also occur when there is a high standard of living. In the absence of other sources of income, reduction of the wage rate must in the somewhat longer run lead to a position of negative elasticity of the supply of labor.

Voluntary unemployment might also be considered to exist as a special case of the situation of positive elasticity when unions declare a strike to achieve a higher wage rate.

Involuntary unemployment, on the other hand, will occur when workers are prepared to work but cannot find work because the demand for labor is insufficient. If the labor market has the form which has been assumed by Walras and his followers as typical of all markets, this type of unemployment can occur in the equilibrium position in one case only. According to

Walras, the market operates in such a way that each supplier and each demander assumes the price as given. In the equilibrium position the price is such that the quantity demanded equals the quantity supplied. There would then be work for everybody who offers his services for work. There would be only one exception to this reasoning, namely, if the supply curve were horizontal at the equilibrium point. This would mean that a large number of workers would be prepared to work just at the current wage rate but not at a somewhat lower rate (see Fig. 43). It would then be possible that, at the current wage rate, demand would be smaller than the total number of workers who offer their services, whereas at the lower wage rate supply would be smaller than demand.

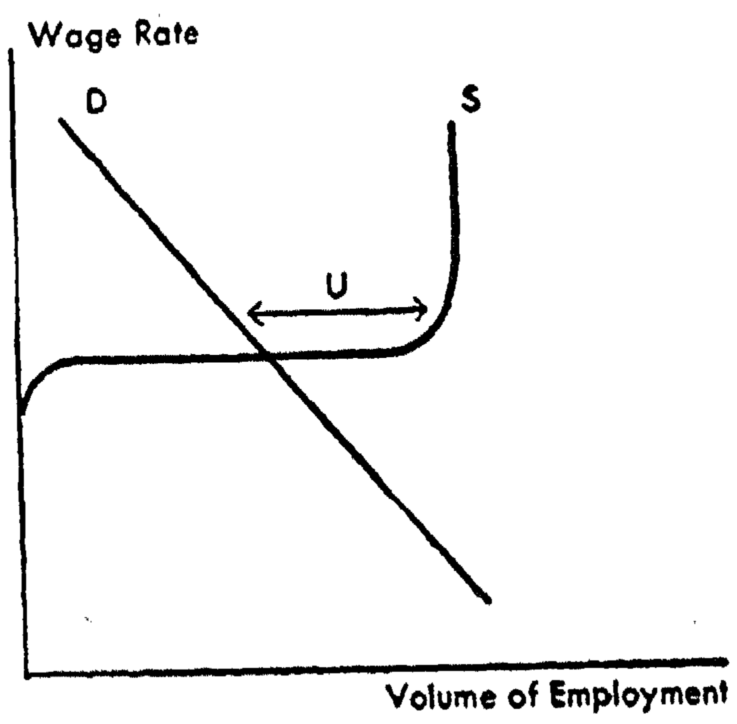

FIG. 43.—Equilibrium in the labor market if the supply curve has a horizontal part (u = unemployment).

This situation may also be described in another way. One may say that the type of market is not the one generally assumed by Walras but rather the following. In the labor market, as in many other markets, the suppliers determine the price at which they are prepared to supply; the buyers determine the quantity they are prepared to take at that price. On this assumption it is quite possible that the total supply, in this particular case the total supply of labor, will actually not be absorbed.

Within the framework of the Walrasian assumptions concerning the organization of the market, it would still theoretically be possible that there was no point of equilibrium, because the demand curve would have no point of intersection with the supply curve. In practice, however, this will probably amount to the same as the preceding case, because in reality there will always be a price, if only a price determined by the preceding period.

Involuntary unemployment is therefore possible at a relatively low level of demand. As we have mentioned, such a situation can occur, and in all probability did occur after 1918 in a number of countries, as a result of a low level of the autonomous demand factors.

MONEY

So far we have abstracted from the "veil of money." In order to approximate reality further, we must take money into account. Now, therefore, we picture all transactions as paid in money: the worker receives his wage, the capitalist his interest, in money; both use this money for the purchase of commodities. The factors of production and the product itself now have a price in terms of money, and the real wage rate to which we have referred so far is represented by the ratio between the money wage rate and the money price of the product. If, in fact, everybody were guided by real prices, nothing would be changed in the physical sphere. All relations and tendencies would remain unchanged. It would be necessary only to add a monetary sphere to our picture. This monetary sphere would be simplest in character if the economic subjects considered money only as a means of exchange and a unit of account and not as an instrument of saving, that is, a means for the accumulation of wealth. They would then hold money only for what Keynes has called the "transactions motive," that is, to meet payments they have to make before new money income becomes available. For any given level of money prices a fixed amount of cash will be needed; if prices are twice as high, double the quantity of money will be required. Conversely, if the quantity of money is given, the level of money prices will adjust itself to this quantity. Assuming no changes in the physical sphere, prices will change in proportion to changes in the quantity of money. If, further, owing to peculiarities of banking legislation (itself based on the beliefs of the economic subjects concerning the requirements of a sound monetary system), the quantity of money is proportional to the gold stock, the price level will be proportional to the latter at constant levels of production. If there are changes in the volume of output, the level of money prices will be proportional to the ratio of the gold stock over the volume of turnover. Studies by Cassel have shown that, in particular, the long waves in the price level satisfy approximately this relationship.

The picture becomes somewhat more complicated if money is used not only as a means of payment but also as a form of accumulation of wealth. There will then be a tendency to keep part of one's capital in the form of cash because cash has the advantage of liquidity. The greater the importance attached to liquidity, the larger the amount of cash held for the liquidity motive and hence not used for purchases of commodities. If the liquidity preference increases, the money price level will tend to fall. An increase in the liquidity preference will occur, for instance, in times of great uncertainty. The best-known example of such a tendency to increase liquidity occurred during the great depression of 1932.

The amount of money held will depend also on the rate of interest. By keeping money in the form of cash, one foregoes the possibility of investing it at interest; a possible interest return is sacrificed. The higher the rate of interest, the less will be the preparedness to accept this sacrifice. The level of money prices will depend, accordingly, also on the rate of interest.

Additional complications will occur if the economic subjects are guided in their decisions concerning the real sphere not only by real prices but also by money prices. In that case the "money illusion" makes itself felt. This will occur, for instance, if the satisfaction of the worker depends to some extent on the level of his nominal wage. It would lead us too far to follow the consequences of the money illusion. With respect to long-run movements, it has to be borne in mind that this illusion is usually only a temporary phenomenon, based on the fact that the subjective valuation of money is based on a previous level of prices. Essentially, the money illusion is an error which has only a second-order effect on the longer-run developments. For short-run movements its consequences may be much more important, but these are not the subject of this chapter. We must break off here our discussions of the long-run development, particularly because statistical observation has not proceeded far enough to permit more than establishing the very general tendencies which we have discussed so far.