CAMBRIDGE

Cambridge University Press
978-0-521-07629-6 - The Neoclassical Theory of Production and Distribution
C. E. Ferguson
Excerpt
More information

1

INTRODUCTION

1.1 Scope

The object of this book is to present a systematic and thorough statement of the neoclassical theory of production and distribution. A decade or so ago this would not have required much space; and whatever the final product, it could only have been an embellishment of Carlson and Shephard, or of Samuelson's beautifully concise 'Comprehensive Restatement of the Theory of Cost and Production.'[1] To be sure, one would have had to add something on the microeconomic theory of distribution; but this could have been easily done by relying upon the works of Hicks, Stigler, Chamberlin, and others.[2]

But matters have changed dramatically within the past few years. Emphasis and professional interest have shifted significantly. Now most research centers on macroeconomic theory, on *aggregate* production functions and their implications for aggregate input substitution, distribution and technological progress (even though there now seems to be a trend toward microeconomics again). Since the behavior of these aggregates has a material effect upon the national economy, there has been a concomitant rise in the interest attached to econometric studies of production, distribution, and technological progress.

In a sense the chronical above is parochial because it concentrates upon developments whose origins are found predominantly on this side of the Atlantic.[3] A group of English economists, centered around Cambridge, has taken an altogether different view of the macroeconomic aspects of distribution and technological progress. While there are certain differences within the group, the tenor of thought is reflected in the works of Kaldor.[4] He argues that it is impossible, or at least meaningless, to distinguish

[1] Carlson [1939], Shephard [1953], and Samuelson [1947, pp. 57–89].
[2] Hicks [1932], Stigler [1939], Stigler [1946], and Chamberlin [1936].
[3] Without serious overstatement, I think one can say that the *origins* are to be found largely in the works of Samuelson and Solow.
[4] Kaldor [1955], [1957], [1959], [1961], and Kaldor and Mirrlees [1962]. See especially Kaldor [1959, pp. 220–4]. I have quoted Kaldor because he is somewhat more outspoken than the rest. Mrs Robinson has doubtless done the most significant work in this vein of theory.

between shifts of a production function and movements along it (i.e. the substitution of capital for labor). In his words,'. . .just as technical progress causes accumulation, the process of accumulation stimulates the growth of knowledge and know-how. Hence it is useless to analyze the effects of capital accumulation in terms of a production function which assumes a given state of knowledge.' [1959, p. 221.] Instead, Kaldor thinks it more sensible to take as given, not the relation between capital stock[1] and output that underlies the production function, but the relation between the rate of accumulation and the rate of change of output.

Upon this foundation several Cambridge economists have built macroeconomic theories purporting to explain capital accumulation, technical progress, and distribution.[2] These theories occupy an important place in the literature and have given rise to no small amount of debate. Hence a coverage of these developments is also necessary if the claim of thoroughness is to be made. But the line is drawn at this point. Other macroeconomic theories of distribution are ignored, and the reader is referred to Davidson's excellent account of them.[3]

Throughout the study primary emphasis is attached to neoclassical theory and, within this framework, to microeconomic theory. Thus Part I is devoted to the microeconomic theories of production and distribution. The (neoclassical) macroeconomic theory of production and distribution is the subject of Part II. The concluding chapters, 15 and 16, are devoted to an exposition and critique of some 'alternative' theories of aggregate distribution and technical change.

1.1.1 Neoclassical theory[4]

The terms 'neoclassical theory' and 'macroeconomic theory' have so far been used without regard to precise definition. In a way this does no great damage because they are well established in our jargon. Yet these terms enter significantly throughout the book; so we must pause for exact description.

'Neoclassical theory' has different implications to different people; but on two counts there would seem to be a uniformity of opinion. These concern the ways in which 'neoclassical theory' differs from 'classical

[1] Kaldor's terminology is used here. As we shall presently see, concentrating upon the flow of capital services, the relevant variable for the production function, allows one to incorporate changes in the state of knowledge and know-how in the production function.

[2] The reader may think that *Cambridge* has been overemphasized. However, on this score see the explicit statement of a leading member of the group, Pasinetti [1962, p. 267].

[3] Davidson [1960].

[4] For a similar treatment, see Ferguson [1965*c*]. In writing this section I have had the advantage of comments by Professor J. J. Spengler.

theory', a term I shall conveniently leave undefined except to say it refers to the world of Malthus and Ricardo. First, neoclassical theory is based upon the assumption that there are no fixed, nonaugmentable factors of production. This contrasts sharply with the classical assumption of a fixed supply of land. Second, in neoclassical theory the rate of growth of population or of the labor force is assumed to be determined exogenously.[1] This too contrasts sharply with classical theory, in which population is very much an economic variable.

Beyond these points there is less accord. To some, neoclassical theory implies that the production function is smoothly continuous and at least twice differentiable. Others add the stipulation that marginal products are positive and continuously diminishing. Neither of these requirements is imposed in Part I. Indeed the subject matter of chapters 2 and 3 is production with fixed technological coefficients; and in chapters 4 and 5 the 'uneconomic regions' of production are discussed. However, when we come to macroeconomics in Part II, the aggregate production function is usually assumed to be smoothly continuous and twice differentiable, with first partial derivatives always positive and second partial derivatives always negative.

1.1.2 Macroeconomic theory[2]

The type of macroeconomic theory discussed here needs two comments, the first of which may be disposed of quickly. So far as this book is concerned, macroeconomic theory deals with *real* aggregate magnitudes; the nominal price level is neither a variable nor a parameter in any model presented. It might seem more appropriate to label this type of analysis 'general equilibrium theory'; and I would raise no objections because the macroeconomic theory of this book is squarely, if somewhat tenuously, based upon general equilibrium theory.

The last statement leads to the second point. General equilibrium theory is essentially microeconomic in character, dealing with the price-output decisions of individual businessmen and the price-purchase decisions of households. Aggregates appear only when they are well defined and when the process of aggregation does not impose restrictive conditions. Indeed, aggregates usually appear only as market balance equations or industry equilibrium conditions. The logical unity of general equilibrium theory is

[1] This statement does not imply that the rate of participation of the labor force is exogenously determined. However (see 1.4.3 below), we are concerned only with full employment situations; the cyclical behavior of output and employment is not under consideration.

[2] A discussion along these lines may be found in Ferguson [1964*b*]. Also see Kuenne [1963, pp. 22–39] and Solow [1959].

beautiful; and once one has worked his way through the complete model, he may stand back and admire its beauty. But that is about all he can do. Some simplifying assumptions are necessary if the theory is to be rendered usable.

There are at least three avenues of approach. First, one may choose to adopt Marshall's approach: to ignore the general interdependence of the economic system, to impound most variables in a *ceteris paribus* assumption, and to analyze the behavior of individual economic units in this carefully circumscribed environment. This is, in fact, the approach adopted in Part I of this book.

Second, one may reduce the general equilibrium system to manageable proportions by making some drastically simplifying (linearizing) assumpions concerning behavioral or technological relations that appear in the model. This is the method of Leontief; and it seems to be powerful indeed for a certain set of problems, mainly of a short-run nature. It is not, however, the set of problems upon which our attention is focused.

Finally, one may aggregate. One way is to define special aggregates and special techniques of aggregation, so that the underlying microeconomic relations entail the existence of corresponding macroeconomic relations. The rigor of general equilibrium theory is thus preserved; but two objections may be raised. First, this method, even if practicable, is not practical.[1] The second objection has been demonstrated forcefully by Grunfeld and Griliches [1960]. Their principal argument is that *in practice* we do not know enough about the underlying microeconomic relations to specify them perfectly. Thus each micro-equation will contain specification errors, and these errors tend to be magnified when aggregated. Consequently, they conclude that a direct aggregate '...equation may explain the aggregate data better than all micro-equations combined if our micro-equations are not "perfect". Since perfection is unlikely,... aggregation is not necessarily bad if one is interested in the aggregates.'[2]

Alternatively, the macroeconomic system may be constructed *by analogy* with the corresponding microeconomic system. Thus one may assume, for example, that the aggregate economic system possesses an aggregate production function, that the system behaves as though it minimizes the cost of producing a stipulated output, and that the system rewards each input according to competitive imputation. In constructing macroeconomic systems by analogy one sacrifices the rigor of general equilibrium theory in favor of empirical feasibility. Index numbers and conventionally defined aggregates are used in lieu of the specially defined aggregates required for precise aggregation. Whether this method is successful or not is itself an

[1] For abundant evidence on this score, see Theil [1954].
[2] Grunfeld and Griliches [1960, p. 10].

empirical question whose answer is treated very lightly in this book. But the macroeconomic theory discussed is the macroeconomic theory constructed by analogy with the corresponding microeconomic theory.

1.2 Method and level of presentation

The pure theory of production and distribution is our exclusive concern; and little or no effort is made to show the relevance of the analysis to issues of current economic policy. Theory is the concern of the theoretician, and to him the book is directed. In no way is this a practical man's guide to production; the method is strictly that of model analysis of model situations. What Dewey has said of his own book applies to this as well: '...this book employs the method of austere, sustained, and, I regret, largely humorless abstraction that has served economics so well in the past. Given the excruciating complexity of so many of the problems encountered..., I cannot see that any other method will allow us to cut through to first principles and deal with these problems according to their importance.'[1]

The level of presentation varies quite widely. Some chapters or parts of chapters are elementary indeed; other passages are relatively advanced. The differences in level stem from my desire to present a thorough and comprehensive statement of the theory under consideration. The method of presentation varies as well. For the most part, mathematical models of production and distribution are discussed; and except for Part II, no effort is usually made to reduce the number of variables under consideration. Yet whenever it seemed helpful, I have not eschewed graphical analysis, numerical examples, or drastic simplification by reduction in the number of variables.

The subject matter of this book is taken from the public domain of economic literature, hopefully with citation. The published works of many economists comprise the sources from which the material was developed. The intended merit of the book lies in the systematic exposition and analysis of the theory and in the explanation and interpretation of some of its more difficult aspects. Nonetheless, portions of chapters 2, 3, 6, 9, and 11–14 are believed to contain original material.

1.3 Definitions

The jargon of production theory is well established and broadly understood. Nonetheless, some possible ambiguity might be removed if careful definitions are given to a few terms.

[1] Dewey [1965, p. vii].

1.3.1 Inputs and outputs

For our purposes an *output* or *product* is any good or service whose fabrication or creation requires one or more scarce resources. When concrete examples are cited, they refer exclusively to goods output simply because it is easier to conceive of the precise set of resources used. Furthermore, one important dimension of products is ignored throughout the book; in particular, our discussion is generally applicable only to that set of outputs whose production does not involve 'time' or 'storage' in an essential way. Thus, for example, the production processes that convert new wine into vintage wine and May wheat into January-consumable wheat are not included. To do so would take us rather far afield into the dynamics of speculation. Still worse, it would make our dangerous skirting of the theory of capital and growth even more perilous.

Next, a *factor of production* or an *input* is defined as any scarce resource[1] used in the production of a good or service. According to the paragraph above, one class of inputs is excluded, namely 'time' or 'storage'. The remainder may conveniently be divided into two broad groups. First, there is a set of inputs, called *ingredient inputs*, that are used up or consumed in the process of production. At times, especially in Part II, we assume that ingredient inputs must enter the production process in fixed proportions one to the other. The recipe, or 'blueprint' in modern terminology, is given and cannot be changed without altering the nature of the product. Since this class of inputs is assumed to enter in fixed proportions, one is able to subtract its cost from the market value of the commodity and thereby determine the 'value added' by those inputs whose proportions are not fixed.

The second class is composed of those inputs that render a flow of services, which are 'consumed' in the production process; but the inputs themselves are not consumed. Capital and labor are the principal, and certainly the most interesting, inputs belonging to this set. In Part II our attention is focused upon these two inputs, which are usually assumed to be homogeneous aggregates.

1.3.2 The short run and the long

For many years economists have indulged in a bit of circularity that has been rewarding. More specifically, it has become customary to define the *short run* as that period of time during which the quantity of one or more inputs cannot be changed. The quantity of input or of its maximum

[1] Using the term 'scarce resource' in these two paragraphs is, in a sense, begging an important economic question. However, it is a question whose answer demands a full model of general equilibrium, rather than a model of the output side alone.

flow of services is fixed and can be neither augmented nor diminished. Such an input is called a *fixed input* in the short run. The *long run* is then defined as that period of time (or planning horizon) in which all inputs are variable, none fixed. The circularity enters because a fixed input is defined as one whose quantity or flow of services cannot be changed in the short run.

It is sometimes argued that no inputs are ever fixed and, consequently, that one cannot distinguish between short and long runs. Whether or not certain inputs are instantaneously augmentable or not seems to be a moot question, which fortunately does not require an answer. No input must, of necessity, be fixed to permit one to take a partial derivative or a cross-section of a production surface. The cross-section may be analyzed for one or more values of the variables conceptually held constant. In what follows the study of these cross-sections is referred to as short-run analysis. It is recognized that a 'short run' might not exist; yet to act as though it exists creates a convenient analytical fiction that is fully justified by the mathematical processes used to define it.

1.4 Assumptions

A large number of assumptions are employed throughout the book. Most of these are specific to the topic immediately under consideration; others apply uniformly throughout (with the exception of Part II, chs. 15, 16). Whenever particular assumptions arise, e.g. the assumption of either perfect or imperfect competion in the input or output markets, they are explicitly noted. However, there are two pervasive assumptions of such importance as to merit discussion at the outset.

1.4.1 The production function

A production function, which may be specified in a mathematical or tabular form, is assumed to exist for each good or service. According to conventional definition,[1] a production function shows the *maximum* output attainable from any specified set of inputs, i.e. any set of quantities of ingredient inputs and flows of services of other inputs. In general, no further limitations are imposed except that the set of outputs and inputs must be nonnegative. Finally, the production function is a single-valued mapping from input space into output space inasmuch as the maximum attainable output for any stipulated set of inputs is unique.

Two broad classes of production functions are discussed in Part I, *fixed-* and *variable-proportions production functions*. A production process is

[1] See, for example, Samuelson [1947, p. 57]. For a somewhat different view, see Smithies [1935, pp. 122–9].

characterized by fixed proportions if, and only if, each level of output technologically requires a unique combination of inputs. If the technologically determined input–output ratio is independent of the scale of production for each input, the production process is characterized by *fixed input coefficients*. In this case, the fixed-proportions production function is homogeneous of degree one. If the input–output ratios are not independent of scale, but if all pairs of input ratios are constant, the fixed proportions production function is homogeneous. However, it is not homogeneous of degree one throughout; in this case, the degree of homogeneity may be constant or it may change with the scale of operation.

One more variant of the fixed-proportions production function is conceivable: the pairs of input ratios are technologically fixed for each level of output, but they change as output changes. This specification is, of course, inconsistent with homogeneity of degree one. It is not discussed in what follows because it would seem to imply that some of the inputs change qualitatively when the scale of production changes. If such cases must be discussed, it would seem more reasonable to suppose that there are several different production functions, each of which is relevant within a certain range of output.

A variable-proportions production function is one in which the same level of output may be produced by two or more combinations of inputs. For convenience, we always assume that variable-proportions functions are smoothly continuous. This assumption restricts us to a subset of the full class of variable-proportion functions, and it rules out boundary solutions.[1] However, the essential *technological* feature of variable proportions is preserved: one input may be substituted for another while maintaining a constant level of output. Of course, as we shall see, the production coefficients are in fact 'fixed' by economic considerations, given the set of ratios of input prices. At efficiency points, observable economic behavior is much the same regardless of the class of functions. Yet there is an important technological difference, and it is important to preserve it.

1.4.2 Fixed and variable proportions, limitational and limitative inputs

At this point it is convenient to pause for an additional definition or classification of inputs. The origin of this classification is presumably Frisch [1931]; but it has been refined and elaborated by Georgescu–Roegen.[2] His formulation is adopted here, especially as presented in

[1] Any member of the excluded group, in which there is only a finite number of input combinations that will yield the same level of output, may be viewed as a set of fixed-proportion production functions. This approach is briefly treated in Part I, chs. 2–3.

[2] Georgescu-Roegen [1935], [1951], [1955], and [1960]. On this point, also see Schneider [1934] and Zassenhaus [1935].

[1955, pp. 299–302]. The definitions are illustrated graphically by figures 1 and 2 under the assumption that there are only two inputs. These figures show two isoquants, labeled Q_0 and Q_1, derived from the following production functions:

$$Q = \min\left(\frac{x_1}{\alpha}, \frac{x_2}{\beta}\right), \tag{1.4.1}$$

and

$$Q = f(x_1, x_2), \tag{1.4.2}$$

where x_1 and x_2 are the amounts of inputs X_1 and X_2, α and β are constants, and f is continuous and everywhere differentiable.

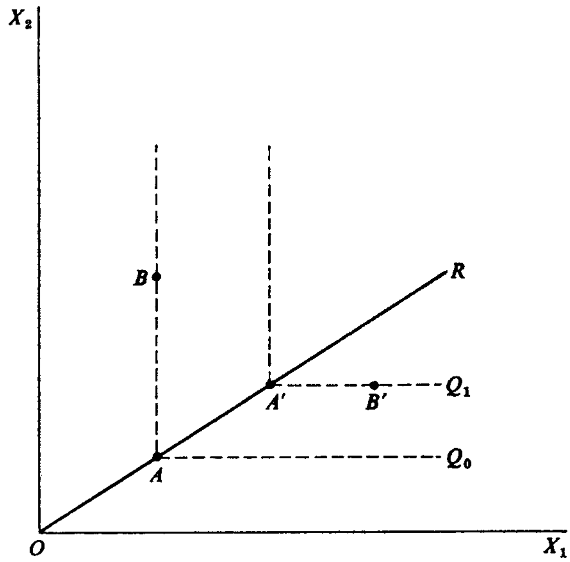

Fig. 1

Now for the definitions. An input is said to be *limitational* if an increase in its usage is a necessary, but not sufficient, condition for an increase in output. According to this definition, both X_1 and X_2 are limitational at points A and A' in figure 1. An increase in either input, the usage of the other remaining constant, would not expand output. Thus increased usage of either input alone is only a necessary condition. It should be noted that no points such as A or A' appear in figure 2. In other words there are no limitational inputs if the production function is characterized by variable proportions.[1] Limitationality imposes a special structure upon the pro-

[1] When ingredient inputs enter the production function as fixed coefficients, while other inputs have technologically variable coefficients, the production function is a mixture of fixed and variable proportions. In this case, the ingredient inputs may become limitational at certain points. This situation, which is treated briefly in Part I, chs. 4 and 5, has been analyzed by Smithies [1936].

duction function, namely, for every level of output, there is a factor combination such that all inputs are limitational. If the limitational production function is also homogeneous, as assumed above, it is completely defined by a single process. This single process is represented by the ray *OR* in figure 1.

Limitativeness, on the other hand, is a characteristic of every production function. An input is said to be *limitative* if an increase in its usage is both a necessary and a sufficient condition for an increase in output. Every pro-

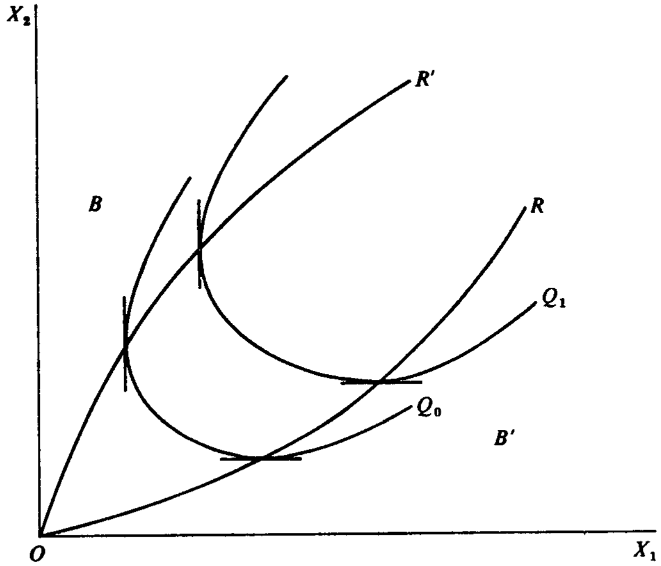

Fig. 2

duction function has a domain of limitativeness for each input. In figure 1, factor X_1 is limitative at point *B* and at every point on the vertical portions of the isoquants. Similarly, X_2 is limitative at point *B'* and at every other point on the horizontal segments of the isoquants.

In figure 2 the domains or regions of limitativeness are more pronounced. In the open region B (the open curvilinear cone X_2OR'), X_1 is limitative. In the region *B'* (the open curvilinear cone X_1OR), X_2 is limitative. On the other hand, X_2 is superfluous in region *B* and X_1 is in region *B'*. Similarly, X_2 is superfluous at *B* in figure 1 and X_1 is at *B'*. More generally, in figure 1, X_2 is superfluous within the open linear cone X_2OR and X_1 is superfluous within the open linear cone X_1OR. Along the common boundary *OR*, both inputs are limitational rather than limitative. This points to another technological characteristic that distinguishes fixed- from variable-proportions production functions.