RICHARD R. NELSON AND SIDNEY G. WINTER

AN EVOLUTIONARY THEORY OF ECONOMIC CHANGE

THE BELKNAP PRESS OF HARVARD UNIVERSITY PRESS CAMBRIDGE, MASSACHUSETTS, AND LONDON, ENGLAND

8

Neoclassical Growth Theory: A Critique

MUCH OF THE ECONOMIC THEORY developed by the great classical economists was concerned with exploring patterns of long-run economic change. Their thinking was strongly influenced by their recognition of technological advance and capital formation as important aspects of the historical transformations they witnessed. Whereas Ricardo and Malthus were (in some respects) pessimistically inclined, Smith before them and most of the classical tradition following them tended to believe that, at least for a considerable time into the future, long-run economic change meant economic progress.

The sharp focusing of microeconomic theorizing on the behavior of firms operating with *given* technologies (in a variety of different market constellations) developed relatively late in the history of economic thought, and came to dominate the textbooks and treatises only after World War II. It is not easy to understand exactly why microeconomic theory was purged of serious concern with long-run change. One reason was that it proved easier to provide a satisfactory mathematical statement of a static theory than a dynamic one. It also seems to have been the case that during the period when these intellectual developments were occurring, economists tended to lose their interest in economic growth, although this may have been a result of the trend that theory was taking, just as much as it was a cause.

In any case, the consequence was that in the 1950s, when many economists again became interested in patterns of long-run economic growth, they found themselves without a well-developed growth theory. First attempts at constructing one appeared in efforts to introduce a more explicit dynamics into Keynesian analysis, through recognition that investment is at once a source of demand for goods and services and a source of increased capacity to produce goods and services. However, the Harrod-Domar growth model, based on an assumption of fixed input coefficients, proved a poor tool for facilitating thinking about rising capital-labor ratios and increasing real incomes per head, which obviously were salient features of observed growth patterns. By the late 1950s, growth theorists had responded to the need to understand these features by borrowing heavily from the intellectual tool kit of static neoclassical microeconomics.

Inevitably, the nature of those neoclassical tools profoundly influenced the approach taken to the explanatory tasks of growth theory. We take it that there is at least rough agreement among economists as to the nature of that task. The minimal set of phenomena to be explained are the time paths of output, inputs, and prices. National economies have grown at various rates over time, and in given eras nations have grown at different rates. Output per worker and capital per worker have grown together. Real wages have risen relative to interest rates. Once one disaggregates the growth experience of particular countries, it is apparent that certain sectors have developed much more rapidly than others and that the sectoral pattern of growth has varied over time. Relative price changes have been correlated with relative productivity growth rates. Although different theories may define and delineate these central phenomena somewhat differently and economists also may divide on questions of the relevance of data of other types (such as productivity differences among firms), almost all economists would agree that a satisfactory theory must be able to explain the above phenomena.

We also take it that most economists would agree that the following are essential elements of the neoclassical explanation.[1] The dominant theme derives from the theory of the firm and production in a competitive industry. At any time, firms are viewed as facing a set of alternatives regarding the inputs and outputs they will procure and produce. Firms choose so as to maximize profits or present value, given the external conditions they face. The economy or sector is assumed to be in equilibrium in the sense that demand and supply are balanced on all relevant markets and no firm can improve its position given what other firms are doing. If we think of a "macro" economy with one sector and with no Keynesian difficulties, growth occurs in the system because over time factors of production expand in supply and production sets are augmented: in an "industry" growth model, changes in demand must be considered as well. The time path of output, input, and prices is interpreted as the path generated by maximizing firms in a moving equilibrium driven by changes in factor demand, factor supply, and technological conditions.

As a glance at Solow's concise survey of growth theory testifies (Solow, 1970), this theory comprises a diverse collection of specific models. The empirical work generated by the theory is similarly diverse. Various neoclassical econometric models have "explained" growth reasonably well on the basis of input growth and technical change, if the criterion is a high R^2. Growth accounting has proceeded apace and has provided an intellectual format for enriching our understanding of the factors that have influenced growth. The theory has been robust in the sense that it continues to survive and to spawn a considerable amount of research that has enhanced our understanding of economic growth. This is a strong plus for neoclassical theory.

However, there is a peculiarity about the success story, which we noted earlier. By the late 1950s it had become apparent that it was impossible to explain very much of the increase in output per worker that had been experienced over the years in developed countries by movements along a production function resulting from increases in capital and other inputs per worker, if constant returns to scale and the other assumptions employed in traditional microeconomic theory were accepted. The "residual" was as large as that portion of total output growth explained by growth of factors of production. For the growth of output per worker, the residual was almost the whole story. The researchers working within the theory found a way to resolve this problem. Earlier, Schumpeter (1934) and Hicks (1932) had proposed that innovation (technical change) could be viewed as a shift in the production function. In the late 1950s Solow's work (1957) made this notion an intellectually respectable part of neoclassical thinking about economic growth. In the empirical work, the residual was simply relabeled "technical advance." Instead of reporting to the profession and the public that the theory explained virtually none of experienced productivity growth, the empirical researchers reported their "finding" that technical change was responsible for 80 (or 85 or 75) percent of experienced productivity growth.

1. The Residual Explanation of Economic Growth

Technological Change as a Residual "Neutrino"

This type of intellectual sleight of hand is not peculiar to economic analysis, and reasonable toleration for it is not necessarily inimical

1, Much of the following discussion was first presented in Nelson and Winter (1973). The analysis of growth accounting follows Nelson (1973).

to the progress of science. The neutrino is a famous example in physics of a "labeling" of an error term that proved fruitful. Physicists ultimately found neutrinos, and the properties they turned out to have were consistent with preservation of the basic theory as amended by acknowledgement of the existence of neutrinos. A major portion of the research by economists on processes of economic growth since the late 1950s has been concerned with more accurately identifying and measuring the residual called "technical change," and better specifying how phenomena related to technical advance fit into growth theory more generally. The issue in question is the success of this work.

Considerable effort has gone into developing the concept of technical change within a production function framework and into modifying that framework to make technical change endogenous to the neoclassical system rather than exogenous. Regarding the first part of the task, the effort can be viewed as augmenting the specification of the production function so as to include more terms—for example, a term that can be interpreted as "total factor productivity" or terms that can be interpreted as the "efficiency" of labor or of capital. These terms are then treated as variables, not constants, within the system. Technical advance is brought into the standard neoclassical format for economic behavior by postulating that these terms are a function of past investments (in activities called research and development) aimed specifically to advance them. The standard profit maximization hypothesis has been employed regarding these investments.

A variety of empirical studies have proceeded guided by the above conceptual structure, and have come up with conclusions that are qualitatively consistent with it. For example, if one assumes that the profitability of an invention is proportional to the sales of an industry, one would expect that changes over time in the amount of inventing directed toward different industries would be correlated with changes in the sizes of industries, and that at any moment in time there would be more inventing going on relevant to "large" industries than to small ones. These are exactly Schmookler's conclusions, based on his use of patents as an indicator of inventing (Schmookler, 1966).

A special version of the theory focuses on technical advance to "save" or increase the efficiency of various factors of production used in producing a particular product. In this version of the theory a rise in the price of one factor relative to another should, other things equal, lead to an increase in efforts aimed to augment the efficiency of that factor relative to others. Recent work by Hayami and Ruttan (1971) and others, directed toward agriculture, shows that both time series and cross-country data are roughly consistent with that theory.

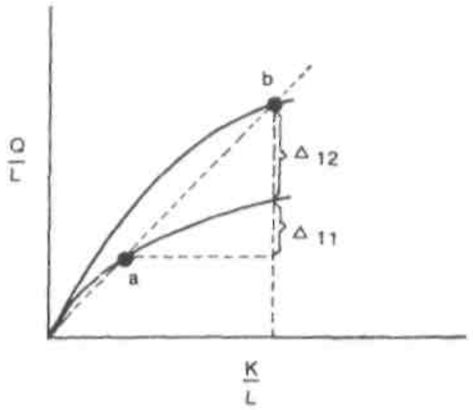
The Identification Problem

In the case of the neutrino, the characteristics of the unobserved particle were relatively well pinned down by prevailing theory (assuming that the theory itself was viable). In the case of technical change, neoclassical theory did not specify very well how "large" or important technological change must be—only that there was "something" there. To see the problem, consider these familiar "stylized facts." Output (gross national product) has been growing at roughly the same rate as capital and at a faster rate than labor; hence, the capital-output ratio has been constant and output per worker and the capital-labor ratio have risen in the same proportion. Factor shares have remained constant; thus, the rate of return on capital has been constant and the wage rate has risen. These "facts" very roughly characterize the Western economic experience that the growth accounting exercises seek to explain.

The facts are inconsistent with an explanation that interprets growth solely in terms of movement along a neoclassical production function. The rise in output per worker would have been less than the rise in the capital-labor ratio, whereas in fact worker productivity has grown at the same rate as capital intensity. And the returns to the factor increasing in relative supply—capital—would have fallen, not remained roughly constant. Thus, the production function must have shifted.

But within the broad framework of interpretation provided by the idea of a shifting production function, there is a wide range of qualitatively different explanations available. Consider the following two, both consistent with the time series data. One is that the underlying production function is Cobb-Douglas (unitary elasticity of substitution) and that technical change has been neutral in the sense of Hicks. The second is that the underlying production function has an elasticity of substitution less than one and that technical change baas been labor-saving. The first interpretation is depicted in Figure 8.1, the second in Figure 8.2. Points *a* and *b* in the two figures are identical and the slopes of the curves (the marginal productivity of capital) at those points also are identical. Thus, both interpretations are consistent with the input, output, and factor price data.

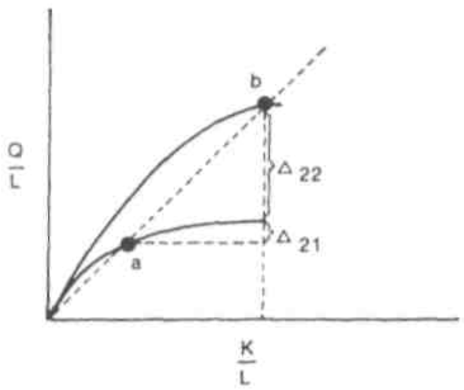
The two interpretations are different in the following "growth accounting" sense. In the case of Figure 8.1, output would have grown by Δ_{11} if capital per worker had grown as it did, but the production function had not shifted. Δ_{12} represents the increase in output per worker not explained by growth of the capital-labor ratio and due, in some sense, to technical change. In Figure 8.2, Δ_{21} can be attributed to growth of capital per worker and Δ_{22} to technical change



8.1 One interpretation of productivity growth.

in the sense above. In the latter interpretation the lower elasticity of substitution means that less of the productivity growth can be attributed to growing capital intensity; hence, more must be attributed to improved technology. Since both interpretations are equally consistent with the time series data, there is no way to choose between them without *a priori* assumptions or other data.

One could view this identification problem as posing difficulties for statistical estimation but as not raising any major theoretical issues; most economists look at the problem this way. For example, it has been proposed that if one had access to cross-section data showing firms operating at the same moment in time using different input coefficients, as well as time series data, one might be able to



8.2 Another interpretation of productivity growth.

disentangle the two sources of growth. Contemporaneous observations would be presumed to reflect the same underlying store of technical knowledge. However, if these firms are within the same economy, these differences in choice of inputs must reflect either the fact that they face different factor prices at the same time, or the fact that they are making different technological choices given the same factor prices; either assumption presents difficulties for the neoclassical formulation that have not really been confronted.

Some Major Conceptual Issues

There are deeper theoretical and conceptual issues behind the scenes. The neoclassical formulation rests on the assumption that at any given time there is a wide range of technological possibilities from which firms may choose, including alternatives that no firm has ever chosen before. The initial period production functions in Figures 8.1 and 8.2 are drawn so as to extend a considerable distance to the right of point *a,* to depict production possibilities employing capital-labor ratios significantly greater than any firm had up to that time experienced. What is the meaning of that? What does one mean when one says that a production possibility exists even though no one is using it or has ever used it? As stated earlier, we do not think it realistic to assume that a sharply defined body of technical knowledge exists that governs production possibilities at input combinations remote from actual experience. Exploration of technologies that have not been used before involves in an essential way the characteristics of "innovation" that we described earlier. If this position is accepted, it is not merely that movements along preexisting production functions explain little of experienced growth. It is that the idea of movements along the production function into previously unexperienced regions—the conceptual core of the neoclassical explanation of growth — must be rejected as a theoretical concept.

The problems with rectifying the production function at remote input combinations are not satisfactorily resolved by grafting onto the theory a neoclassical model of induced innovation. The graft assumes that "inventing" or "doing R&D" is an activity whose out-come can be predicted in advance in fine detail. In effect, there is no difference in the amended theory between moving along the production function by increasing one kind of capital (plant and equipment) through physical investment, and "pushing outward" the production function by increasing another form of capital (knowledge?) through investing in R&D. Both kinds of investments are explained by the same behavioral model. The distinction between innovation and routine operation is totally repressed.

It is repressed at the level of description of the activities involved. There is no room in the neoclassical formulation for nontrivial uncertainty, or for differences of opinion regarding what will work best, or for recognition of the fact that the set of innovation alternatives is shrouded in fundamental ambiguity.

It is repressed in the characterization of the "output" of the activities involved. The models discussed above, which view "shifts" in production functions as resulting from investment undertaken by firms as part of the profit-maximizing portfolio of investments, rest on the presumption that the outcome of research and development is a "private good." Yet certainly there is often an important degree of "publicness" about new knowledge, whether that knowledge is in the form of "blueprints" or in the form of experience. This is so even if the innovating firm tries to restrict access to that knowledge. At the least, knowledge that another firm has done something successfully changes the thinking of other firms regarding what is feasible. And, in some cases, enough knowledge is published or is evident to the sophisticated observer to provide very good clues as to how to proceed.

It could be that the neoclassical induced innovation models implicitly postulate a system of patents. But this certainly is not built specifically into the theoretical formulation. If it were, the theory would need to take account of the fact that firms at any time differ in terms of the technologies they can use, or would have to postulate a perfect system of patent licensing. However, in either case, as long as firms differ in terms of what they come up with as a result of their research and development activities, firms will differ in terms of their profitability.

Inconsistency with Micro Data

The amended neoclassical formulation represses the uncertainty associated with attempts to innovate, the publicness of knowledge associated with the outcomes of these attempts, and the diversity of firm behavior and fortune that is inherent in a world in which innovation is important. Thus, it is unable to come to grips with what is known about technological advance at the level of the individual firm or individual invention, where virtually all studies have shown these aspects to be central. This has caused a curious disjunction in the economic literature on technological advance, with analysis of economic growth at the level of the economy or the sector proceeding with one set of intellectual ideas, and analysis of technological advance at a more micro level proceeding with another.

Over the years economists, other social scientists, and historians have done an enormous amount of research on the more micro aspects of technological change. We shall discuss this literature in some detail in Chapter 11. Suffice it to say here that studies by historians like Landes (1970), Habakkuk (1962), David (1974), and Rosenberg (1972), and by students of industrial organization and technical change like Schmookler (1966), Jewkes, Sawers, and Stillerman (1961), Peck (1962), Griliches (1957), Mansfield (1968), and Freeman (1974) have revealed extremely interesting facts about the technological change process. While some of these are in harmony with neoclassical themes, others are quite discordant. We have, for example, much evidence of the role of insight in the major invention process, and of significant differences in ability of inventors to "see things" that are not obvious to all who are looking. Yet once one has made a breakthrough, others may see how to do similar, perhaps even better, things. The same patterns apparently obtain in innovation. Relatedly, there are considerable differences among firms at anytime in terms of the technology used, productivity, and profitability. Although these studies show clearly that purpose and calculation play an important role, the observed differences among persons and firms are hard to reconcile with simple notions of maximization unless some explicit account is taken of differences in knowledge, maximizing capabilities, and luck. The role of competition seems better characterized in the Schumpeterian terms of competitive advantage gained through innovation or through early adoption of a new product or process than in the equilibrium language of neoclassical theory.

It is not possible to reconcile what is known about the phenomena at a micro level with the intellectual structure used to model technical advance at the macro or sectoral level by arguing that the macro model deals with the average or the modal firm. The differences among firms and the disequilibrium in the system appear to be an essential feature of growth driven by technical change. Neoclassical modeling cannot avail itself of this insight.

There have been a few noteworthy if neglected attempts to square the neoclassical theory of industry production and growth with the observed facts of very considerable diversity of techniques and profitability of firms within an industry at any time. Houthakker (1956) developed a model in which firms at a given time are endowed with different techniques, with each firm being profitable under some sets of product and factor prices but not under others. These techniques are fixed and given, as are the capacities of the various firms. Firms either produce at capacity or produce nothing, depending on the vector of prices. Within such a model it is possible that the aggregate industry data from different periods and different prices will have a form that resembles that of orthodox neoclassical theory. But the model would predict that in a given time there would exist considerable diversity across firms in productivity levels and profitability.

The Houthakker model does not explain why the techniques in existence (with positive capacity) at any time are what they are, and in his model the distribution of capacity over techniques is treated as a constant. There are several different models that "explain" cross-industry diversity of techniques at any time as a result of the dates at which various plants were put in place. See notably Solow, Tobin, von Weizsäcker, and Yaari (1966), Salter (1966), and Johansen (1972). But in these vintage models new investment is always in "best practice" technology, and firms are never uncertain about the characteristics of new technologies. And the evolution of "best practice" is unexplained. Thus, the neoclassical vintage models, at least their present versions, abstract away much of what scholars of the microeconomics of technical advance have learned about the topic.

Theoretical schizophrenia thus forces economists to keep their understandings in different boxes. A central purpose of a theoretical structure—to enable one to see links between apparently disparate phenomena and thus to enable knowledge to be superadditive—is thwarted by this neoclassical partitioning of technical advance. Relatedly, the structure of contemporary formal theory drives a wedge between the analysis of those economists who take the theory seriously, and those; such as economic historians, who pay more attention to the phenomena involved.

The tension has been recognized in the profession. For example, Nordhaus and Tobin have remarked: "The [neoclassical] theory conceals, either in aggregation or in the abstract generality of multisectoral models, all of the drama of events—the rise and fall of products, technologies, and industries, and the accompanying transformation of the spacial and occupational distributions of the population. Many economists agree with the broad outlines of Schumpeter's vision of capitalist development, which is a far cry from growth models made nowadays in either Cambridge, Massachusetts or Cambridge, England. But visions of that kind have yet to be transformed into a theory that can be applied to everyday analytic and empirical work" (Nordhaus and Tobin, 1972, p. 2).

2. The Need for an Evolutionary Approach to Growth Theory

The issue then is this. Following upon the discovery that there was a large "residual" involved in neoclassical explanations of economic growth, and the identification of that residual with technical change, economists undertook a considerable amount of research aimed toward pinning down what technical change actually is. This is just what happened after physicists discovered the neutrino. But what we now know about technical change should not be comforting to an economist who has been holding the hypothesis that technical change can be easily accommodated within an augmented neoclassical model. Nor can the problem here be brushed aside as involving a phenomenon that is "small" relative to those that are well handled by the theory; rather, it relates to a phenomenon that all analysts (or virtually all) acknowledge is the central one in economic growth. The tail now wags the dog. And the dog does not fit the tail very well. The neoclassical approach to growth theory has taken us down a smooth road to a dead end. If an evolutionary approach has advantages as a way of analyzing traditional textbook questions, the arguments for such an approach to growth theory seem overwhelming.

9

An Evolutionary Model of Economic Growth

THE STRENGTHS of the neoclassical approach to economic growth are considerable. Neoclassical theory has provided a way of thinking about the factors behind long-run economic growth in individual sectors and in the economy as a whole. The theoretical structure has called attention to the historical changes in factor proportions and has focused analysis on the relationship between those changes and factor prices. These key insights and the language and formalism associated with them have served effectively to guide and to give coherence to research that has been done by many different economists scattered around the globe. The weakness of the theoretical structure is that it provides a grossly inadequate vehicle for analyzing technical change. In particular, the orthodox formulation offers no possibility of reconciling analyses of growth undertaken at the level of the economy or the sector with what is known about the processes of technical change at the microeconomic level.

The challenge to an evolutionary formulation then is this: it must provide an analysis that at least comes close to matching the power of neoclassical theory to predict and illuminate the macroeconomic patterns of growth. And it must provide a significantly stronger vehicle for analysis of the processes involved in technical change, and in particular enable a fruitful integration of understanding of what goes on at the micro level with what goes on at a more aggregated level.

The key ideas of evolutionary theory have been laid out. Firms at any time are viewed as possessing various capabilities, procedures, and decision rules that determine what they do given external conditions. They also engage in various "search" operations whereby they discover, consider, and evaluate possible changes in their ways of doing things. Firms whose decision rules are profitable, given the market environment, expand; those firms that are unprofitable contract. The market environment surrounding individual firms may be in part endogenous to the behavioral system taken as a whole; for example, product and factor prices may be influenced by the supply of output of the industry and the demand for inputs. In Part III this broad conceptual scheme was incorporated in specific models of selection equilibrium and of the response of firms to changed market conditions. The task now is to devise particular models, consistent with the broad theory, that are especially well suited to analysis of economic growth.

The model presented in this chapter is embodied in a computer simulation program.[1] Simulation techniques have been employed in economic analysis for a variety of different reasons. In some cases (probably comprising the best-known applications) the model is believed to be based on good understanding of a large number of different components of the overall problem. In large-scale macroeconomic models, these may be of the form of estimated behavioral relationships. What is desired is to analyze the effect of various hypothesized changes (the elapse of time, an increase in the tax rate) on a set of variables representing the interactive outcome of a large number of these processes (gross national product, employment, consumption expenditure). The problem is too complicated and constrained, however, to work through analytically. Therefore, the analyst puts the overall model on the computer and "experiments" with the variables whose impact he wants to assess. In cases like this, the analyst has clearly in mind the "structure" of the model he wants to analyze. Although he can analyze a highly simplified form of that model with more conventional techniques, simulation is dictated by an unwillingness to bear the costs of such "oversimplification."

Our situation here is not quite the same. We have some strong qualitative beliefs about a number of components of the model we want to build, but certainly are not rigid about the precise form they should take. We are very flexible about other components, and will choose these so as to enhance the tractability of the model. Our central objective is to build a model that admits, and will likely generate, considerable diversity of behavior at the level of the individual firm. At the same time we want the model to generate aggregative time paths of certain variables, and want to be able to manipulate certain

1. The model and most of the subsequent discussion was presented earlier in Nelson and Winter (1974) and Nelson, Winter, and Schuette (1976).

variables of the model so that these time paths are broadly consistent with historical experience. Also, we want to be able to explore the way in which certain variables defined at the microeconomic level influence these macroeconomic time paths. These requirements naturally lead us to a simulation format.

Needless to say, there are costs involved in working with a simulation rather than an analytic model. For one thing the results are of uncertain generality. If there is a large domain of interesting independent variables and parameters to explore, it is virtually impossible to explore all parts of it. The problem is compounded if the model is stochastic; one is then unsure about the representativeness of the result, even for the parts of the domain explored. In our view, however, the most serious problem with many simulation models is lack of transparency: the models yield results that are not easy to understand. Although this danger is more obvious in simulation than in other modeling strategies, it would be a mistake to believe either that simulation models are inherently opaque or that the results of more traditional analytic techniques are inherently transparent. A random sample of articles from contemporary economics journals is likely to include a substantial proportion of cases in which "conclusions" ground out by traditional analytic techniques take the form of complex mathematical expressions whose substantive economic rationale is extremely difficult (perhaps impossible) to discern.

Also, one can aim for and achieve a considerable amount of transparency in a simulation model by keeping it relatively simple and clean. And this will create opportunities to use simulation and analytic techniques in tandem.

Analysis is, in our view, an important complement of a good simulation study. Special cases of a simulation model (for example, where certain variables are set at zero) may be analytically tractable. It may be possible to construct simple analytic models that capture certain features of the more complicated simulation model; for example, in Chapter 10 we present such a simple analytic model that has much in common with the more complicated simulation model developed here. More generally, simple analytic arguments often can provide an economically meaningful interpretation of the results of simulation experiments.

Simulation, on the other hand, can be a useful adjunct to an analytic approach. Simulation models are not bound by some of the constraints imposed by the requirement for analytic tractability. But the simulation format does impose its own constructive discipline in the modeling of dynamic systems: the program must contain a complete specification of how the system state at $t+1$ depends on that at t and on exogenous factors, or it will not run. In contrast, in orthodox analytic modeling the stress is on equilibrium conditions, and time paths may be treated in an *ad hoc* way or completely ignored.

The opportunity for fruitful exploitation of the complementarity can, however, be largely foreclosed if it is not treated as an important consideration in the design of the simulation model. Most important, the freedom associated with the relaxation of tractability constraints must be exercised with restraint if the output is to be susceptible to analytic checking and interpretation. To introduce complexity in the name of "realism" alone, disregarding the added costs of checking and interpretation, is no more appropriate in the one theoretical endeavor than in the other. It is, in short, a very pernicious doctrine that portrays simulation as a nontheoretical activity, in which the only guiding rule is to "copy" reality as closely as possible. If reality could be "copied" into a computer program, that approach might be productive—but it cannot, and it is not.

1. THE MODEL

An evolutionary model of economic growth must be able to explain the patterns of aggregate outputs, inputs, and factor prices that neoclassical theory "explains." In the exercise here, the standard of reference is provided by Robert Solow's classic article "Technical Change and the Aggregate Production Function" (Solow, 1957). The data addressed in that article comprise gross national product (GNP), capital input, labor input, and factor prices, over a forty-year period. Data beneath these macro aggregates is ignored. Our simulation model must be capable of generating those macro aggregates, but through the route of "building them up" from microeconomic data. And our model must eschew neoclassical analytic components based on well-specified production functions and profit-maximizing behavior and employ in their place the evolutionary theory components of decision rules, search, and selection.

The model involves a number of firms, all producing the same homogeneous product (GNP), by employing two factors: labor and physical capital. In a particular time period, a firm is characterized by the production technique it is using—described by a pair of input coefficients (a_l, a_k)—and its capital stock, K. As in the model presented in Chapter 6, to enable us to exploit the mathematics of finite Markov chains, capital stock is assumed to come in discrete packets. A firm's production decision rule is simply to use all of its capacity to produce output, using its current technique—no slow-down or shut-down decision is allowed for. Thus, at any time, the "state" of a firm can be characterized by a triple (a_l, a_k, K) indexed by time and the identification number for a particular firm. The industry state at time *t* is the (finite) list of firm states at time f. Given the basic behavioral assumption, aggregate output and labor demand are directly determined by the industry state. The wage rate is endogenous, and is determined in each time period by reference to a labor supply curve. The gross returns to capital are simply output (at price equal to one) minus labor payments. Thus, the model can generate or explain the macroeconomic data that Solow addressed.

Changes in the industry state are generated by applying probabilistic transition rules, independently, to the individual firm states. These transition rules result from our specification of search processes and investment rules. In turn, the way we characterize particular transition mechanisms reflects our desire to capture, in stylized form, some of the salient aspects of technical advance and Schumpeterian competition as they have been identified by microeconomic studies. We discuss, first, the transition rules for firms "in business"—that is, with a positive capital stock. Assumptions governing entry will be mentioned later. In the following discussion, a parenthetical delta (5) will identify parameters that have been varied in the experimental runs. The assumptions below, which determine the form of the general model, reveal the kinship of this model with that analyzed in Chapter 6. Yet they differ in important ways.

Technical Change

Use of the term "search" to denote a firm's activities aimed at improving on its current technology invokes the idea of a preexisting set of technological possibilities, with the firm engaged in exploring this set. This connotation seems natural when one is considering R&D aimed to find, say, a seed variety with certain properties or a chemical compound with certain characteristics. It seems less natural when one is considering R&D aimed to develop a new aircraft, or, more generally, R&D activities where the terms "invention" or "design" seem appropriate. Instead of exploring a set of preexisting possibilities, R&D is more naturally viewed in these contexts as creating something that did not exist before. And surely modern research on hybrid seeds and pharmaceuticals involves creating as much as discovering.

But for the purposes of our evolutionary modeling, the distinction here is one of semantics not substance. The R&D activities of our firms will be modeled in terms of a probability distribution for coming up with different new techniques. We will discuss this in terms of sampling from a distribution of existing techniques. But alternatively we could discuss it in terms of a distribution of things that a firm might "create." In either case, that distribution might be a function of time (opportunities might evolve over time), a firm's R&D policy (some firms might spend more or perform different kinds of R&D than others), the firm's existing technique (search may be largely local), and other variables.

In the particular model explored in this chapter, time *per se* is not an element; there is a given set of techniques to be found; a firm's R&D "policy" is modeled as involving "satisficing." And what a firm comes up with as a result of its R&D is much influenced by its prevailing technique and the prevailing techniques of other firms.

Satisficing. To highlight the similarity of the model employed here to the equilibrium-seeking model of Chapter *6,* we assume that if firms are sufficiently profitable they do no "searching" at all. They simply attempt to preserve their existing routines, and are driven to consider alternatives only under the pressure of adversity. Their R&D activity should thus be conceived as representing an *ad hoc* organizational response rather than a continuing policy commitment. This satisficing assumption is a simple and extreme representation of the incentives affecting technical change at the firm level. We dispense with this assumption in the dynamic competition models in Part V, in which the differential profitability of alternative levels of commitment to R&D expenditure is a major focus of concern, but we believe it is adequate for our present purposes. In fact, it seems useful to demonstrate that in an evolutionary model with such conservative firms, there can be continuing innovation in the economy as a whole.

In the simulation runs here, only those firms that make a gross return on their capital less than the target level of 16 percent engage in search. Given that a firm is searching, it either seeks incremental improvements to its present methods or looks to what other firms are doing, but not both at the same time.

Local Search. There is a given constant set of technological possibilities, and each technique is characterized by coefficients a_l and a_k. Technical progress occurs as this set gradually is explored and discovered. For any firm engaging in such exploration, search is "local" in the sense that the probability distribution of what is found is concentrated on techniques close to the current one. The formula used for the distance between techniques *h* and *h'* is

$$D(h, h') = WTL\,|\log a_l^{h} - \log a_l^{h'}| + WTK\,|\log a_k^{h} - \log a_k^{h'}|, \text{ where } WTL + WTK = 1.$$

That is, distance is a weighted average of the absolute differences in the logs of input coefficients. This gives rise to diamond-shaped equal-distance contours in the space of logs of input coefficients. Employment of different values of *WTL* (δ) permits us to treat search with differing degrees of "bias" toward discovering labor- or capital-saving technologies. Probabilities for transitions from a given technique to others are then determined as a decreasing linear function of distance, subject to obvious nonnegativity conditions, an appropriate normalization, and introduction of a probability that no alternative technique will be found. The slope of this linear function is *IN(δ)*, where *IN* stand mnemonically for "ease of *IN*novation." The larger (less negative) the value of *IN,* the more likely it is that the search process will uncover technologies with input coefficients significantly different from the initial ones.

Imitation. A searching firm may look to what other firms are doing. If it does, the probability that it will find a particular technique is proportional to the fraction of total industry output produced by that technique in the period in question. Alternatively we might have assumed that imitation is focused on "best practice," and we do so in models presented later. The assumption here is more consonant with models of diffusion, where what is best practice is not obvious to a firm *ex ante* but where widely used techniques attract attention.

The actual probabilities of "finding" different techniques for a firm that is searching are, then, a weighted average of the probabilities defined by "local search" and the probabilities defined by "imitation." The relative weights on local search and imitation are characterized by the parameter *IM* (δ), where *IM* is a mnemonic for "emphasis on *IM*itation." A high value of *IM* denotes a regime where search is more likely to be over what other firms are doing and less likely to be of the "local search" type than it would be in regimes where the value of *IM* is low.

An alternative rule turned up by the search process is adopted by the firm only if it promises to yield a higher return, per unit capital, than the firm's current rule. (Since the firm's capital stock is independently determined, the return-per-unit-capital criterion gives the same result as a test based on anticipated total profit.) The wage rate employed in this comparison is the one associated with the current industry state. There is an element of random error in the comparison: the capital and labor input coefficients employed in the test are not the true values for the alternative technique, but the products of the true values and realizations of independent normal deviates. A firm in business misjudges the input coefficient of an alternative technique by an amount that exceeds 20 percent about a third of the time.

Investment

Our characterization of the determinants of changes in the sizes of firms can be described much more compactly. The capital stock of a firm with positive capital in the current state is first reduced by a random depreciation mechanism; each unit of capital is, independently, subject to a failure probability of D = 0.04 each period. The capital stock, thus reduced, is then increased by the firm's gross investment in the period. Gross investment is determined by gross profit, where gross profit πK is revenue Q minus wage bill WL minus required dividends RK. (More precisely, gross investment is gross profit rounded to the nearest integer, the rounding being necessary because capital stock is integer-valued and gross profit is not.) This rule is applied even when gross profit is negative, subject only to the condition that the resulting capital stock not be negative. The higher the value of *R* (δ), the smaller the investment the firm is able to finance.

Entry

As indicated above, we make special assumptions about entry. A firm with zero capital in the current state is a potential entrant and "contemplates" the use of a production decision rule. If its decision rule implies a gross rate of return to capital in excess of 16 percent calculated at current prices, it becomes an actual entrant with probability 0.25. If it does enter, its capital stock is determined by a draw on a distribution that is uniform over the integers from five to ten. (Entry is relatively infrequent, and the contribution it makes to gross investment is minor when averaged over several periods.) Other firms (those contemplating rules that do not meet the rate-of-return test) remain at capital stock zero with probability one. The assumptions about search by potential entrants differ slightly from the assumptions about search by firms already in the industry; these will be mentioned when needed.

The Labor Market

The price of labor is endogenous to the model, being determined by the exogenous supply and endogenous demand for labor. The prevailing wage rate influences the profitability of each firm, given the technique it is using, and, in turn, the behavior of the industry as a whole is a powerful, but not unique, influence on the wage rate. The simulation program admits all wage determination equations of the form

$$w = a + b\left(\frac{L_t}{(1+g)^t}\right)^c,$$

where t is the time period, L, is the aggregate labor use in the period, and *a, b, c,* and *g* are constants. When $g = 0$, labor supply conditions are constant over time, and the model as a whole is a Markov process with constant transition probabilities. A nonzero *g* corresponds to changing labor supply conditions; the model as a whole remains a Markov process, but with time-dependent transition probabilities.

The Markov process defined by the above relations may be summarized as follows. At any moment the capital stocks of extant firms, together with their techniques, determine their required labor inputs and their outputs. Industry output and total labor employment then are determined. Total labor employment determines the industry wage rate. Given the wage rate, the gross profitability of each firm is determined.

Firms that make a gross rate of return of less than the target level engage in search. Of those firms that are searching, some attempt to innovate and others to imitate the techniques used by more profitable firms. Firms screen the techniques that they have uncovered by search, and if they deem them more profitable they are adopted and the old ones discarded. Firms that had been earning more than the target level, or that do not come up with techniques they deem better than the ones they had, keep their old techniques.

Extant firms invest in the purchases of new capital the earnings they have left after paying wages and required dividends. Their net investment equals gross investment minus depreciation. New firms may enter the industry at positive capital stock if the profitability of the technique they were contemplating exceeds the target level.

Thus, the next-period techniques of all firms are determined (probabilistically), and so are the next-period capital stocks. The "industry state" for the next period then has been established.

Calibration

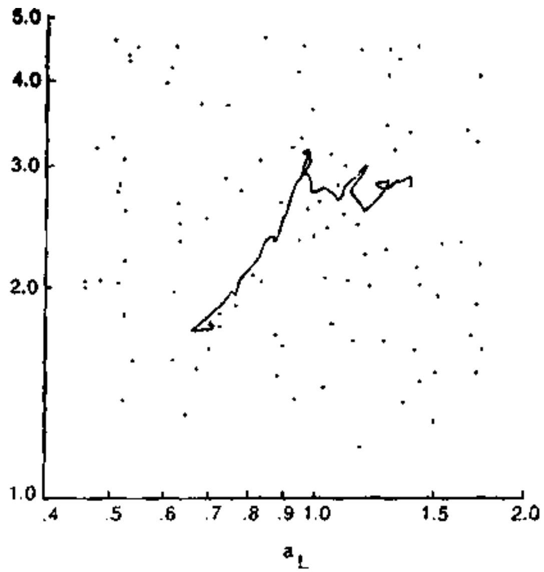
The model will generate a time path of firm and industry inputs and output, and a time path of the industry wage rate and firm and industry rates of the return on capital, the labor share, and the capital share. One central question we are exploring is whether a model of the sort described above is capable of generating time paths of the macroeconomic variables that are similar to the actual observed time paths of these variables (in particular to those displayed in the data analyzed by Solow). The initial conditions of the model were set so that they roughly-corresponded to the conditions revealed in Solow's data for 1909.[2] Thus, we initially endowed our firms with techniques that, on average, had roughly the input coefficients displayed by the Solow data for 1909. We assigned an initial amount of capital to each firm and positioned the labor supply curve so that, given the implied labor requirements for the initial period, the wage rate equaled the 1909 wage rate and the initial capital-labor ratio roughly matched the 1909 data. (For reasons of convenience we chose an initial total capital stock of three hundred units.) Given that wage rate and the choice of input coefficients, the initial average rate of return on capital of our firm must be roughly equal to that in the Solow data for 1909. And labor's and capital's shares of income under initial conditions of the model also will be consonant with the actual Solow data for 1909. The data analyzed by Solow also determined the set of possible techniques (input coefficient pairs) built into the model. The techniques were determined by random choice from the uniform distribution over a square region in the space of logarithms of input coefficients.[3] The region includes, with room to spare, all of the historical coefficients implied by Solow's data. We judged that distinguishing one hundred possible techniques in this region would permit adequate representation of cross-sectional diversity and historical change. This scatter is displayed in Figure 9.1, along with the actual time paths of input coefficients from the Solow data. An important question being explored is whether the (average) input coefficients of our simulation model can be induced to display a time path that is similar to the actual one.

The time path of the input coefficients, and of related variables like the capital-labor ratio, obviously will depend on how labor and capital grow over time in the model. Given the broad specification of the model's logic, this will depend on the particular parameter settings of some of the key variables. Thus, in the runs reported here we have assumed that the labor supply curve shifted to the right over the period of time at a rate of 1.25 percent per year. This is roughly consistent with the observed historical rate.

2. More precisely, the attempt was made to set initial values so that period 5 of the simulation run would approximately agree with the 1909 values.

3. A slight compromise of the random choice procedure was made: the scatter chosen was one of four generated, and it was selected because it was most free of "holes"—areas of the square in which no techniques occurred.

4.



9.1 Input coefficient pairs for unit output, with Solow's historical input coefficient values.

Varying Some Key Parameter Values

One question we are asking about the model is whether, under plausible parameter settings of the sort described above, it can generate time paths of the macroeconomic variables comparable to those actually observed. Another range of issues being explored involves the connections between variables defined at the microeconomic level and the macroeconomic time path.

The "localness of innovative search" assumption built into the model implies that, at a microeconomic level, most innovations are relatively minor. It is possible, however, to vary the localness of search—to make it easier (more likely) or harder (less likely) for a firm to discover a technique significantly different (far away) from the one it has. (Specifically, the relevant parameter here is the slope of the relationship between the distance of an alternative technique from the current one and the probability that the alternative will be discovered.) If search is less local, if major innovation is easier, a firm is more likely to come up with innovation that is markedly inferior. But given that its profitability checks are reasonably reliable, it will not adopt such innovations. On the other hand, the innovations that it does adopt are likely on average to be bigger (involving a larger de-cline in the input coefficients). To what extent would the ease of major innovation, in the above microeconomic sense, show up in, say, a faster rate of growth of labor productivity or of total factor productivity? One's faith in the model's ability to represent micro-macro links would be severely strained unless there were some such association. By choosing different settings of the "ease of major innovation" parameter, it is possible to explore this question.

It also is possible to vary the parameter that determines what fraction of a firm's "searching" will be directed to what other firms are doing, rather than toward possible innovations. What differences would this make? The logic of the system at the micro level would suggest that if more search is directed toward imitating and less toward innovating, the production techniques of firms will tend to be bound together more closely. The competitive race would be "closer." And one implication of this might be that firms tend to remain together in size, as well as in technology. By calculating some measure of industry concentration at the beginning and end of the simulation runs, one can explore the effect of different degrees of emphasis on imitation on the extent to which concentration evolves over time in the model economy, and perhaps on some other variables. If interesting and plausible connections show up in the simulation results, these might form hypotheses to be tested against real-world data.

One can also vary the required dividend rate. If the dividend payout is low, the rate of growth of the capital stock ought to be higher than it would be if the payout were higher. This higher capital stock might be expected to lead to higher labor demand, thus to higher wages and to a tendency to adopt less labor-intensive techniques, when the cost of capital is low than it would when the cost is high. Another influence on the evolution of the capital-labor ratio might come from the extent to which search is easier in a capital-saving direction or in a labor-saving direction. One can vary this within the model by choosing different weights on the distance measure regarding innovative search.

In our simulation models we employed two different settings for each of the variables discussed above: the ease of major innovation, the emphasis on imitation, the cost of capital, and the labor-saving bias of search. That is, we undertook runs (of fifty periods each) with sixteen different sets of parameter settings. The sixteen runs comprise all possible combinations of levels of the four experimental factors, with two levels for each factor,

All of the experimental runs were initiated with the same assignments of techniques to thirty-five firms. In the eight runs with a high dividend payout rate, the fifteen firms in business each had twenty units of capital. In the runs with low required dividends, firms in business each had twenty-two units of capital. These initial capital values were chosen to put the system in approximate "equilibrium"— that is, with roughly zero expected net investment in the initial period. To have started all runs at the same industry state, ignoring the implications of the different parameter values, would have been a straightforward but naive approach to the problem of achieving "identical" initial conditions for the different runs. Drastic differences in the aggregate outcomes in the early periods would then have been implied by the differences in parameter values; no such strong effects are visible in the results as they stand.

2. The Growth Record of the Simulated Economy

The computer output describing the experimental simulation runs contains abundant quantitative detail and is rich in qualitative patterns. Firms thrive and decline; new techniques appear, dominate the scene briefly, and then fade away. Time series for most aggregate data display strong trends, and also a good deal of short-period fluctuation. The stack of paper containing the description of the total of eight hundred years of synthetic history is over eight inches high. It is clear that it must be summarized fairly drastically for the purpose of this discussion.

How do the aggregative time series look? In a word, plausible. In Table 9.1 the results of one simulation run and the real data addressed by Solow are displayed side by side. There is, of course, no reason to expect agreement between the real and simulated data on a year-to-year basis. The simulation run necessarily reflects nonhis-

Table 9.1. Selected time series from one simulation run, compared with Solow data, 1909–1949.

	Q/L[a]		K/L[b]		W[c]		S_k[d]		A[e]	
Year	Sim	Solow	Sim	Solow	Sim	Solow	Sim	Solow	Sim	Solow
1909	0.66	0.73	1.85	2.06	0.51	0.49	0.23	0.34	1.000	1.000
1910	0.68	0.72	1.84	2.10	0.54	0.48	0.21	0.33	1.020	0.983
1911	0.69	0.76	1.83	2.17	0.52	0.50	0.25	0.34	1.040	1.021
1912	0.71	0.76	1.91	2.21	0.50	0.51	0.30	0.33	1.059	1.023
1913	0.74	0.80	1.94	2.23	0.51	0.53	0.31	0.33	1.096	1.064
1914	0.72	0.80	1.86	2.20	0.61	0.54	0.15	0.33	1.087	1.071
1915	0.74	0.78	1.89	2.26	0.56	0.51	0.24	0.34	1.108	1.041
1916	0.76	0.82	1.89	2.34	0.60	0.53	0.21	0.36	1.136	1.076
1917	0.78	0.80	1.93	2.21	0.59	0.50	0.23	0.37	1.159	1.065
1918	0.78	0.85	1.90	2.22	0.62	0.56	0.21	0.34	1.169	1.142
1919	0.80	0.90	1.96	2.47	0.57	0.58	0.29	0.35	1.190	1.157
1920	0.80	0.84	1.94	2.58	0.64	0.58	0.19	0.32	1.192	1.069
1921	0.81	0.90	2.00	2.55	0.61	0.57	0.25	0.37	1.208	1.146
1922	0.83	0.92	2.02	2.49	0.65	0.61	0.21	0.34	1.225	1.183
1923	0.83	0.95	1.97	2.61	0.70	0.63	0.17	0.34	1.243	1.196
1924	0.86	0.98	2.06	2.74	0.64	0.66	0.26	0.33	1.274	1.215
1925	0.89	1.02	2.19	2.81	0.59	0.68	0.33	0.34	1.293	1.254
1926	0.87	1.02	2.07	2.87	0.74	0.68	0.15	0.33	1.288	1.241
1927	0.90	1.02	2.16	2.93	0.67	0.69	0.25	0.32	1.324	1.235
1928	0.91	1.02	2.18	3.02	0.70	0.68	0.23	0.34	1.336	1.226
1929	0.94	1.05	2.27	3.06	0.68	0.70	0.28	0.33	1.370	1.251
1930	0.98	1.03	2.47	3.30	0.62	0.67	0.37	0.35	1.394	1.197
1931	0.99	1.06	2.46	3.33	0.70	0.71	0.29	0.33	1.408	1.226
1932	1.02	1.03	2.57	3.28	0.69	0.62	0.32	0.40	1.435	1.198
1933	1.02	1.02	2.46	3.10	0.85	0.65	0.16	0.36	1.452	1.211
1934	1.04	1.08	2.45	3.00	0.85	0.70	0.19	0.36	1.488	1.298
1935	1.05	1.10	2.44	2.87	0.87	0.72	0.17	0.35	1.500	1.349
1936	1.06	1.15	2.51	2.72	0.82	0.74	0.22	0.36	1.499	1.429
1937	1.06	1.14	2.55	2.71	0.83	0.75	0.22	0.34	1.500	1.415
1938	1.11	1.17	2.74	2.78	0.76	0.78	0.32	0.33	1.543	1.445
1939	1.10	1.21	2.66	2.66	0.88	0.79	0.20	0.35	1.540	1.514
1940	1.13	1.27	2.75	2.63	0.84	0.82	0.25	0.36	1.576	1.590
1941	1.16	1.31	2.77	2.58	0.90	0.82	0.23	0.38	1.618	1.660
1942	1.18	1.33	2.78	2.64	0.95	0.86	0.20	0.36	1.641	1.665
1943	1.19	1.38	2.79	2.62	0.93	0.91	0.22	0.34	1.652	1.733
1944	1.20	1.48	2.80	2.63	0.97	0.99	0.20	0.33	1.672	1.856
1945	1.21	1.52	2.82	2.66	0.97	1.04	0.20	0.31	1.683	1.895
1946	1.23	1.42	2.88	2.50	0.96	0.98	0.22	0.31	1.694	1.812
1947	1.23	1.40	2.89	2.50	0.98	0.94	0.21	0.33	1.701	1.781
1948	1.23	1.43	2.87	2.55	1.01	0.96	0.18	0.33	1.698	1.809
1949	1.23	1.49	2.82	2.70	1.04	1.01	0.15	0.33	1.703	1.852

a. Q/L = Output (1929 dollars per man-hour; Solow data adjusted from 1939 to 1929 dollars by multiplying by 1.171, the ratio of implicit price deflators for GNP).
b. K/L = Capital (1929 dollars per man-hour).
c. W = Wage rate (1929 dollars per man-hour; Solow data adjusted from 1939 to 1929 dollars).
d. S_k = Capital share (equals one minus the labor share).
e. A = Solow technology index. (Recalculation on the basis of figures in other columns will not check exactly, because of rounding of those figures. Solow figures shown for 1944–49 are correct; the values originally published were in error.)
Note: These data are from run 0001; see Table 9.2 for key to run numbering.

torical random influences. But more than that, and of particular importance to this comparison, the simulation model, unlike Solow's analysis of the real data, generates its own input history on the basis of very simple assumptions about behavior and institutional structure. The real period in question involved eposides of economic depression and war, and while these episodes might be considered as historical random events, the simulation model is not prepared to deal with them realistically. The same trend in the labor force, the same Say's Law assumption, the same link of investment to retained earnings persist year by year. Since the model's historical accuracy is so sharply limited by these considerations we have not attempted to locate parameter settings that would, in any sense, maximize similarity to the real time series. For example, it would have been easy to assure a better match of initial conditions.

Rather, the question we think should be addressed is whether a behavioral-evolutionary model of the economic growth process, of the sort described in the preceding section, is capable of generating (and hence of explaining) macro time series data of roughly the sort actually observed. So considered, we regard the simulation as quite successful. The historically observed trends in the output-labor ratio, the capital-labor ratio, and the wage rate are all visible in the simulated data. The column headed *A* in the table shows the Solow-type index of technology, computed on the contrafactual assumption that the simulated time series were generated by a neutrally shifting neo-classical production function. The simulated average rate of change in this measure is about the same as in the Solow data (indicating, essentially, that we have chosen an appropriate value in this run for our localness-of-search parameter). It is interesting to note, however, that our simulated world of diverse simple-minded firms searching myopically in a continuing disequilibrium generates a somewhat smoother aggregate "technical progress" than that found by Solow in the real data for the United States. For example, our series shows only five incidences of negative technical progress, whereas Solow's series shows eleven—and the run shown is typical in this respect.

Table 9.2 presents data on each run for each of several variables, observed at period forty of the run.[4] Also displayed are the corresponding figures, where these exist, for the thirty-sixth period (1944) and the fortieth period (1948) of the Solow data. Given the experimental design, it is convenient to distinguish the runs by numbering them in the binary system. The interpretive key to this numbering is explained in the note to Table 9.2.

It is plain that the simulation model does generate "technical progress" with rising output per worker, a rising wage rate, and a rising capital-labor ratio, and a roughly constant rate of return on

Table 9.2. Values of aggregative variables at period forty.

Run	$\frac{K}{L}(40)^a$	$A(40)^b$	$a_L(40)^c$	$a_k(40)^d$	$C_4(40)^e$	$\Delta w(40)^f$	$\Delta Q(40)^g$
0000	2.796	1.727	0.832	2.326	0.560	1.4	3.6
0001	3.129	2.391	0.592	1.851	0.521	2.5	5.0
0010	2.519	1.712	0.846	2.131	0.383	1.6	3.4
0011	4.242	2.716	0.477	2.025	0.387	3.2	3.6
0100	2.035	1.855	0.825	1.678	0.645	1.8	3.8
0101	2.695	2.106	0.679	1.829	0.404	2.4	4.5
0110	2.686	1.658	0.841	2.258	0.405	1.4	6.0
0111	2.703	2.123	0.672	1.817	0.388	2.1	4.6
1000	3.015	1.746	0.800	2.411	0.476	2.1	4.7
1001	4.511	2.359	0.524	2.364	0.457	2.4	5.6
1010	4.332	2.098	0.600	2.599	0.443	1.9	4.4
1011	4.258	2.450	0.514	2.190	0.325	2.8	4.3
1100	3.212	1.835	0.705	2.265	0.491	1.9	4.3
1101	3.391	2.190	0.600	2.034	0.518	2.6	5.1
1110	3.031	1.963	0.705	2.136	0.394	1.9	5.3
1111	3.315	1.913	0.682	2.260	0.327	1.9	4.1
Solow (1944)	2.63	1.856	0.675	1.776	—	—	—
Solow (1948)	2.55	1.810	0.699	1.784	—	1.7	—

a. K/L = Capital-labor ratio.

b. A = Solow technology index. (Solow figures for 1944 and 1948 are correct; the values originally published were in error.)

c. a_L = Average labor input coefficient, L/Q.

d. a_k = Average capital input coefficient, K/Q.

e. C_4 = Four-firm concentration ratio. (Initial value = 0.206.)

f. Δw = Rate of change of wages, percent per period.

g. ΔQ = Rate of change of output, percent per period.

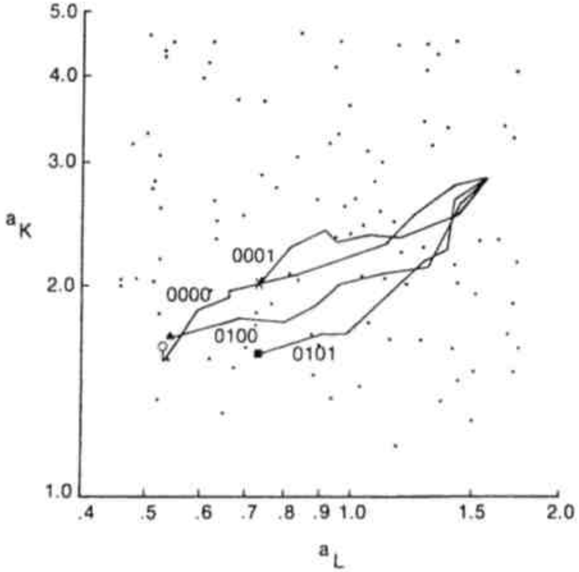
Note: Runs are numbered in binary, $X_{WT}X_RX_{IM}X_{IN}$. When $X_{IN} = 0$, the probability of discovery of a technique declines with distance with slope −6.0; when $X_{IN} = 1$, the slope is −4.5. In the $X_{IM} = 0$ setting, search activity involves imitation with probability .2 for extant firms; when $X_{IM} = 1$, that probability is .4. When $X_R = 0$, the required dividend rate $R = .02$; when $X_R = 1$, $R = .06$. With the $X_{WT} = 0$ setting, there is no bias in search, whereas when $X_{WT} = 1$, $WTL = .4$ and $WTK = .6$.

4. The reason for focusing on values observed late in the run is to allow plenty **of** time for the different parameter settings to display their distinctive influences on the industry state. The reason for observing at period 40 rather than, say, at period 50 is that a few of the runs display, in the late periods, clear "boundary effects" associated with proximity of average input coefficients to the edge of the **region from which the** decision rules were chosen.

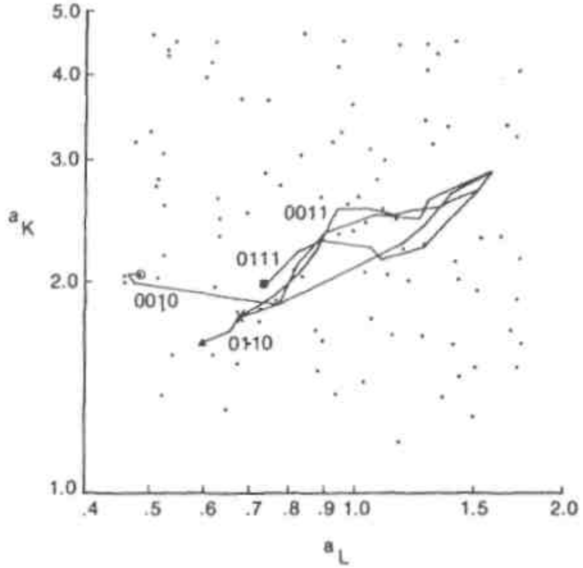
capital. The rates of change produced correspond roughly to those in the Solow data. Also, some individual runs produce values quite close to the Solow values for the variables measured — for example, runs 0101 and 0111.

Figures 9.2-9.5 display the time paths of the average input coefficients generated by the sixteen runs. To keep the figures relatively uncluttered, the values are plotted for the initial period and at periods 5, 10, and so on thereafter. In Figure 9.6 the input coefficient track for one run (1110) is compared with the track implied in the Solow data. The case shown is one in which there is close agreement at the initial point, and also forty periods later, but there is a wide divergence in between. The divergence is associated with the fact that, while the simulated track gives the impression of taking a relatively constant direction, there is a sharp turn in the track of the real data, suggestive of a change in the underlying regime. The apparent break occurs between 1929 and 1934. Perhaps it would be asking too much of the simulation model, committed as it is to full employment, to reproduce that break.

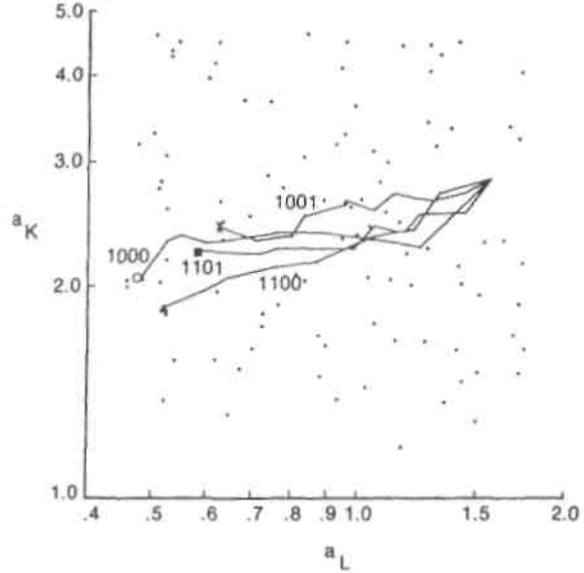
It seems interesting to ask: If a neoclassical economist believed the aggregative time saving generated by the simulation model to be real data, and tested his theory against the data, what would he conclude?



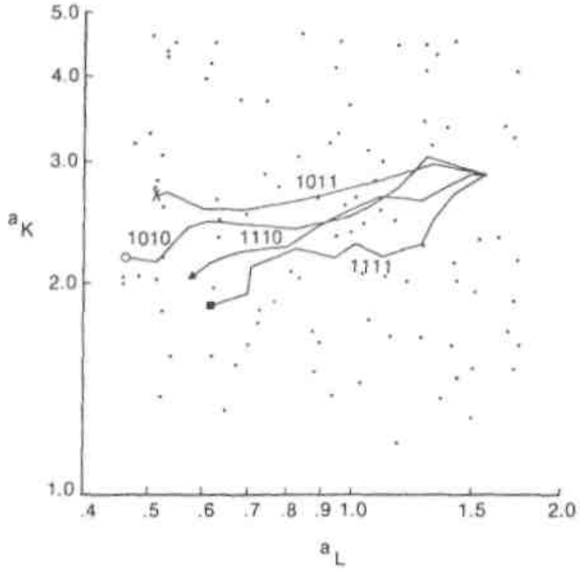
9.2 Average input coefficient paths for four runs with low emphasis on imitation and no bias in search.



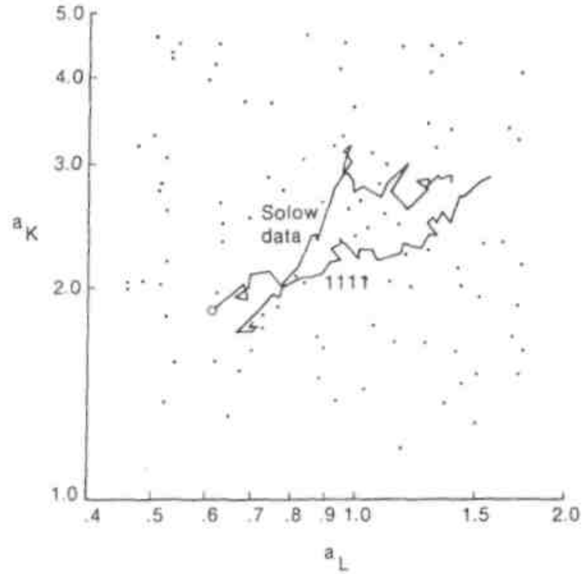
9.3 Average input coefficient paths for four runs with high emphasis on imitation and no bias in search.



9.4 Average input coefficient paths for four runs with low emphasis on imitation and labor-saving bias in search.



9.5 Average input coefficient paths for four runs with high emphasis on imitation and labor-saving bias in search.



9.6 Average input coefficient paths: Solow data compared with run 1111.

Table 9.3. Cobb-Douglas regressions, Solow method.

$$\log\left(\frac{Q(t)/A(t)}{L(t)}\right) = a + b \log\left(\frac{K(t)}{L(t)}\right)$$

Run	b	R^2	Run	b	R^2
0000	0.195	0.993	1000	0.211	0.968
0001	0.184	0.990	1001	0.268	0.991
0010	0.244	0.996	1010	0.261	0.994
0011	0.214	0.993	1011	0.256	0.986
0100	0.219	0.985	1100	0.325	0.999
0101	0.248	0.988	1101	0.241	0.987
0110	0.301	0.998	1110	0.249	0.978
0111	0.193	0.942	1111	0.313	0.997

The answer depends on the particular simulation run from which the data are taken and on the particular test. But by and large it seems that he would believe that his theory had performed well. (Of course, if he also looked at the microeconomic data and observed the inter-firm dispersion of techniques and differential growth rates, he might ponder a bit whether his theory really characterized what was going on. But the pondering would likely conclude with the consoling thought that macro theories need not square with micro observations.)

Tables 9.3 and 9.4 display the results of fitting Cobb-Douglas production functions, by each of two methods, to the aggregate time

Table 9.4. Cobb-Douglas regressions with time trend.

$$\log Q(t) = a + b_1 \log K(t) + b_2 \log L(t) + b_3 t$$

Run	b_1	b_2	b_3	R^2	Run	b_1	b_2	b_3	R^2
0000	0.336	0.649	0.012	0.999	1000	0.505	0.550	0.008	0.998
0001	0.681	0.541	0.011	0.999	1001	0.648	0.360	0.011	0.999
0010	0.201	0.764	0.016	0.998	1010	0.723	0.336	0.009	0.999
0011	0.728	0.158	0.017	0.997	1011	0.532	0.505	0.015	0.998
0100	0.281	0.654	0.016	0.999	1100	0.637	0.444	0.008	0.999
0101	0.222	0.833	0.017	0.999	1101	0.669	0.448	0.010	0.999
0110	0.405	0.593	0.009	0.998	1110	0.479	0.545	0.013	0.999
0111	0.075	0.658	0.013	0.999	1111	0.641	0.547	0.007	0.998

series data for each experimental run. The Solow procedure was followed in generating Table 9.3. The percentage neutral shift in the hypothetical aggregate production function was calculated in each period, and the technology index *A(t)* constructed. The index was then employed to purge the output data of technological change, and the log of adjusted output per labor unit was regressed on the log of capital per labor unit. The observations were taken from periods 5-45 of the simulation run, to give us a sample size the same as Solow's and to minimize possible initial-phase and terminal-phase effects on the outcomes. The regressions in Table 9.4 are based on an assumed exponential time trend in the technology index and involve the logs of the absolute magnitudes rather than ratios to labor input. The same sample period was employed.

The most noteworthy feature of these results is that the fits obtained in most of the cases are excellent: half of the R^2 values in Table 9.3 exceed 0.99, and more than half of those in Table 9.4 equal 0.999. The fact that there is no production function in the simulated economy is clearly no barrier to a high degree of success in using such a function to describe the aggregate series it generates. It is true that the fits obtained by Solow and others with real data are at least as good as most of ours, but we doubt that anyone would want to rest a case for the aggregate production function on what happens in the third or fourth decimal place of R^2. Rather, this particular contest between rival explanatory schemes should be regarded as essentially a tie, and other evidence consulted in an effort to decide the issue.

Thus, a model based on evolutionary theory is quite capable of generating aggregate time series with characteristics corresponding to those of economic growth in the United States. It is not reasonable to dismiss an evolutionary theory on the grounds that it fails to provide a coherent explanation of these macrophenomena. And the explanation has a certain transparency. As we discussed earlier, many of the familiar mechanisms of the neoclassical explanation have a place in the evolutionary framework.

Consider, for example, the empirically observed nexus of rising wage rates, rising capital intensity, and increasing output per worker. Our simulation model generated data of this sort. In that model, as in the typical neoclassical one, rising wage rates provide signals that move individual firms in a capital-intensive direction. As was proposed in Chapter 7, when firms check the profitability of alternative techniques that their search processes uncover, a higher wage rate will cause to fail the "more profitable" test certain techniques that would have "passed" at a lower wage rate, and will enable to pass the test others that would have failed at a lower wage rate. The former will be capital-intensive relative to the latter. Thus, a higher wage rate nudges firms to move in a capital-intensive direction compared with that in which they would have gone. Also, the effect of a higher wage rate is to make all technologies less profitable (assuming, as in our model, a constant cost of capital), but the cost increase is proportionately greatest for those that display a low capital-labor ratio; thus, a rise in wages tends to increase industry capital intensity relative to what would have been obtained. And output per worker will be increased; a more capital-intensive technology cannot be more profitable than a less capital-intensive one unless output per worker is higher.

While the explanation here has a neoclassical ring, it is not based on neoclassical premises. Although the firms in our simulation model respond to profitability signals in making technique charges and investment decisions, they are not maximizing profits. Their behavior could be rationalized equally well (or poorly) as pursuit of the quiet life (since they relax when they are doing well, and typically make only small changes of technique when they do change) or of corporate growth (since they maximize investment subject to a payout constraint). Neither does our model portray the economy as being in equilibrium. At any given time, there exists considerable diversity in techniques used and in realized rates of return. The observed constellations of inputs and outputs cannot be regarded as optimal in the Paretian sense: there are always better techniques not being used because they have not yet been found and always laggard firms using technologies less economical than current best practice.

On our reading, at least, the neoclassical interpretation of long-run productivity change is sharply different from our own. It is based on a clean distinction between "moving along" an existing production function and shifting to a new one. In the evolutionary theory, substitution of the "search and selection" metaphor for the maximization and equilibrium metaphor, plus the assumption of the basic improvability of procedures, blurs the notion of a production function. In the simulation model discussed above, there was no production function—only a set of physically possible activities. The production function did not emerge from that set because it was not assumed that a particular subset of the possible techniques would be "known" at each particular time. The exploration of the set was treated as a historical, incremental process in which nonmarket information flows among firms played a major role and in which firms really "know" only one technique at a time.

We argue—as others have before us—that the sharp "growth accounting" split made within the neoclassical paradigm is bothersome empirically and conceptually. Consider, for example, whether it is meaningful to assess the relative contribution of greater mechanization versus new technology in increasing productivity in the textile industries during the Industrial Revolution, of scale economies versus technical change in enhancing productivity in the generation of electric power, or of greater fertilizer usage versus new seed varieties in the increased yields associated with the Green Revolution. In the Textile Revolution the major inventions were ways of substituting capital for labor, induced by a situation of growing labor scarcity. It could plausibly be argued that in the electric power case, various well-known physical laws implied that the larger the scale for which a plant was designed, the lower the cost per unit of output it should have. However, to exploit these latent possibilities required a considerable amount of engineering and design work, which became profitable only when the constellation of demand made large-scale units plausible. Plant biologists had long known that certain kinds of seed varieties were able to thrive with large quantities of fertilizers and that others were not. However, until fertilizer prices fell, it was not worthwhile to invest significant resources in trying to find these varieties. In all of these cases, patterns of demand and supply were evolving to make profitable different factor proportions or scales. But the production set was not well defined in the appropriate direction from existing practice. It had to be explored and created.

We argued in Part II that at any given time the set of techniques that an individual can control skillfully, or that an organization can control routinely, likely does not extend very far beyond those that are being more or less regularly exercised. Relatedly, we proposed that an attempt to employ a technique significantly different from those likely involves a nontrivial amount of deliberation, research, trial and feedback, and innovation. But in Chapters 6 and 7, and here again, models in which only a small part of changed input-output relations could be regarded as "routine" (moving along a production function) displayed patterns over time that had many of the qualitative properties of movements along the production functions of orthodox theory. The model in this chapter is somewhat extreme in endowing a firm with only one technique that it can operate routinely at any time. It would not be inconsistent with evolutionary theory to assume that a firm at any time is capable of operating a small number of alternative techniques, with various decision rules employed to determine the mix. In this case a larger share of factor substitution in response to changing prices would have been accounted for by along-the-rule movements. But it is interesting that even with along-the-rule responses excluded completely, an evolutionary model is capable of generating, and hence explaining, data that orthodox theory explains only by recourse to the unrealistic assumption that firms have large, well-defined production sets that extend well beyond the experienced range of operation.

The question of the nature of "search" processes would appear to be among the most important for those trying to understand economic growth, and the evolutionary theory has the advantage of posing the question explicitly. In the simulation model, we assumed that technical progress was the result strictly of the behavior of firms in the "sector" and that discovery was relatively even over time. However, it is apparent that the invention possibilities and search costs for firms in particular sectors change as a result of forces exogenous to the sector. Academic and governmental research certainly changed the search prospects for firms in the electronics and drug industries, as well as for aircraft and seed producers. In the simulation, the "topography" of new technologies was relatively even over time.[5] However, various studies have shown that often new opportunities open up in clusters. A basic new kind of technology becomes possible as a result of research outside the sector. After a firm finds, develops, and adopts a version of the new technology, a subsequent round of marginal improvements becomes possible. This appears to be the pattern, for example, in the petroleum-refining equipment and aircraft industries. However, this pattern does not show up in the manufacture of cotton textiles (after the Industrial Revolution) or in the automobile industry, where technical advance seems to have been less discrete. The search and problem-solving orientation of an evolutionary theory naturally leads the analyst to be aware of these differences and to try somehow to explain or at least characterize them.

The perspective on the role of the "competitive environment" is also radically different in the evolutionary theory, and leads one to focus on a set of questions concerning the intertwining of competition, profit, and investment within a dynamic context. Is the investment of a particular firm strictly bounded by its own current profits? Can firms borrow for expansion? Are there limits on firm size, or costs associated with the speed of expansion? Can new firms enter? How responsive are "consumers" to a better or cheaper product? How long can a firm preserve a technically based monopoly? What kind of institutional barriers or encouragements are there to imitation? The answers to these questions are fundamental to under-

5. Here and subsequently, we use the term "topography" in a metaphorical sense to suggest the role of the cognitive conditions under which the search for new methods takes place. The topography of innovation determines what possibilities can be seen from what vantage points, how hard it is to get from one spot in the space of possibilities to another, and so forth.

standing the workings of the market environment. The specifics of their treatment, like that of the nature and topography of "search," is an empirical issue within our theory.

These kinds of questions can be illuminated by some of the findings of the vast literature on the micro aspects of technological change. Chapter 11 will be concerned specifically with such an exploration. However, some interesting micro-macro links appear in our simulation model.

3. THE EXPERIMENTS

In our discussion above of the logic of the model, we introduced four variables that tie macroeconomic performance to microeconomic behavior and that were varied experimentally in the simulation runs. These variables were the ease of innovation, the emphasis on imitation, the cost of capital, and the labor-saving bias of search. What effect do different settings of these variables have on the macroeconomic time paths in the model?

We adopted a linear regression approach to this question. We considered three different macroeconomic variables: the Solow technology index in year forty, the capital-labor ratio in year forty, and the four-firm concentration index. Our four experimental variables we designated X_{IN}, X_{IM}, X_R, and X_{WT}. We assigned the value one to these variables when (respectively) major innovation was relatively easy, search emphasis was on imitation, the required dividend rate was high, and the search was somewhat biased in a labor-saving direction."

The effects on the period-forty value of the Solow technology index are characterized by the following regression equation:

$$A(40) = 2.335 + \underset{(0.006)}{0.456}\, X_{IN} + \underset{(0.59)}{0.0529}\, X_{IM} - \underset{(0.07)}{0.194}\, X_R + \underset{(0.73)}{0.034}\, X_{WT}$$

$$R^2 = 0.705.$$

Figures in parentheses are significance levels. The conjecture that easier major innovation at a microeconomic level should lead to a faster rate of growth of total factor productivity at a macroeconomic level is strongly confirmed. This lends additional confidence that the model provides plausible and understandable connections between the microeconomic phenomena and macroeconomic phenomena of

6. For the explanation of the parameter settings corresponding to the two levels of our experimental factors, see the footnote to Table 9.2.

economic growth. Note that this is not a trivial result, since the rate of growth of total factor productivity and the level of the Solow technology index late in an economy's evolution here are simply macro statistics, and do not correspond directly to features of the model.

Some interesting results also come out of regression analysis of the determinants of the capital-labor ratio in year forty.

$$\frac{K}{L}(40) = 3.353 + \underset{(0.017)}{0.577}\, X_{IN} + \underset{(0.19)}{0.288}\, X_{IM} - \underset{(0.005)}{0.717}\, X_R + \underset{(0.003)}{0.7825}\, X_{WT}$$

$$R^2 = 0.766.$$

The hypothesized effects of factors three and four are strongly confirmed. A higher price of capital, considered as a return that must be paid out and that is not available for reinvestment, does lead to a substantially less capital-intensive mode of production after a period of time. Considered as a growth rate effect, the rise in R from 0.02 to 0.06 produces a decrease of 0.3 percentage points per period in the rate of change of the capital-labor ratio. The effect of the labor-saving search bias introduced by factor four is of comparable magnitude but, of course, in the opposite direction.

The magnitude and significance level of the coefficient of X_{IN} comes as something of a surprise. Why should the capital-labor ratio be higher in a system in which search is less local? On reflection, one possible answer to this question seems to be the following. The general direction of the path traced in input coefficient space does not depend on the localness of search. However, the rate of movement along the path is slower if search is more local. Therefore, given that the path is tending toward higher capital-labor ratios (as a consequence of the level chosen for R and the neutrality or labor-saving bias of search), the capital-labor ratio that results after a given number of periods is lower when search is more local. Another possible answer is more Schumpeterian. A high rate of technical progress may produce a high level of (disequilibrium) profits, which in turn are invested. The resulting increase in the demand for labor results in a higher wage and deflects the results of profitability comparisons in the capital-intensive direction. These possible answers are not, of course, mutually exclusive. The regression result regarding concentration is:

$$C_4(40) = 0.495 - \underset{(0.04)}{0.058}\, X_{IN} - \underset{(0.0004)}{0.127}\, X_{IM} + \underset{(0.91)}{0.0028}\, X_R - \underset{(0.22)}{0.033}\, X_{WT}$$

$$R^2 = 0.741.$$

Here, C_4 is the four-firm concentration ratio. The imitation effect is clearly the most pronounced. We have suggested an explanation for this effect in terms of the "closer race." There are actually two distinct mechanisms in the simulation model by which a closer technical race tends to keep concentration down, and both are quite plausible as hypotheses about economic reality. First, as among firms in business, similarity in technique implies similarity in cost conditions, hence in profit rates, and hence in growth rates. Thus, a closer race implies a smaller dispersion of firm growth rates and lower concentration. But, second, potential entrants also stay closer to the technical leaders when imitation is easy and perceived opportunities for profitable entry thus occur more frequently. Since entry tends to occur in a particular (and relatively low) scale range, the amount of capacity added by entry is higher when entry is higher. Considerations of overall industry "equilibrium" imply that the infusion of capacity through entry is partially offset by lower investment by the firms previously in business. Since the latter are typically larger than the entrants, concentration is reduced.

The above analysis of the influences on the concentration of firms is illustrative of a fundamental difference between the neoclassical and evolutionary approaches to growth theory. Neoclassical growth theory is aimed at macro phenomena, and its micro details are instrumental to its macro purposes. Evolutionary theory treats the micro processes as fundamental and treats the macro aggregates as aggregates. Hence, it encompasses a wider range of phenomena; its treatment of the micro details is intended to be subject to test. Thus, for example, we can treat our simulation model not only as an abstract account of the phenomena of aggregate economic growth, but also as an abstract account of the size distribution of firms. This we will do in a later chapter.

4. Summary and Conclusions

We return now to our opening theme. Neoclassical theory has provided a fruitful way of looking at certain macroeconomic patterns of growth. However, it has been strikingly unsuccessful in coming to grips with the phenomena of technological change, and relatedly that theory stands as an obstacle in thinking about microeconomic phenomena and macroeconomic phenomena within the same intellectual frame. In this chapter we have shown that a model based on evolutionary theory can come to grips with the macro phenomena, although at the cost of somewhat greater complexity than that usually involved in neoclassical models. With that increased complexity has come some loss of transparency, although we have argued that the model involves readily discernable relationships between input growth and output growth, and between changes in factor prices and changes in factor proportions. And the gain has been in terms of a characterization of the technological change phenomenon that is much closer to the accounts of those who have studied it carefully, and in terms of the ability to encompass microeconomic phenomena and macro phenomena within the same intellectual framework. *We* have produced an account of economic growth in technical change that is simultaneously consistent (1) in quantitative terms, with the broad features of a certain body of aggregated data; (2) qualitatively, with such phenomena as the existence of cross-sectional dispersions in capital labor ratios and efficiency, and patterns of innovation and diffusion of techniques; and (3) metaphorically, with the empirical literature on firm decision making. These fragments of economic reality (at least) need not be regarded as posing isolated problems to be addressed through special-purpose assumptions. The model's consistency with disparate types of data indicates that it is not merely consistent with the data of any one type, but rather bears a fairly intimate relationship to "what is really going on out there."