Lecture Notes in Economics and Mathematical Systems

Managing Editors: M. Beckmann and W. Krelle

293

D. Batten J. Casti B. Johansson (Eds.)

Economic Evolution and Structural Adjustment

Proceedings, Berkeley, California, USA, 1985

Springer-Verlag

Lecture Notes in Economics and Mathematical Systems

Managing Editors: M. Beckmann and W. Krelle

293

D. Batten J. Casti B. Johansson (Eds.)

Economic Evolution and Structural Adjustment

Proceedings of Invited Sessions on
Economic Evolution and Structural Change
Held at the 5th International Conference
on Mathematical Modelling at the University
of California, Berkeley, California, USA
July 29–31, 1985

Springer-Verlag
Berlin Heidelberg New York London Paris Tokyo

Editors

Prof. Dr. David Batten
Department of Economics, University of Umeå
S-901 87 Umeå, Sweden

Prof. John L. Casti
Institute for Econometrics & Operations Research, Techn. Univ. of Vienna
Argentinierstraße 8, A-1040 Vienna, Austria

Prof. Dr. Börje Johansson
CERUM (Centre for Regional Science Research)
and
Department of Economics, University of Umeå
S-901 87 Umeå, Sweden

ISBN 3-540-18183-0 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-18183-0 Springer-Verlag New York Berlin Heidelberg

Printed in Germany

Printing and binding: Druckhaus Beltz, Hemsbach/Bergstr.
2142/3140-543210

Preface

Since the beginning of the fifties, the ruling paradigm in the discipline of economics has been that of a competitive general equilibrium. Associated dynamic analyses have therefore been preoccupied with the stability of this equilibrium state, corresponding simply to studies of comparative statics. The need to permeate the boundaries of this paradigm in order to open up new pathways for genuine dynamic analysis is now pressing.

The contributions contained in this volume spring from this very ambition. A growing circle of economists have recently been inspired by two distinct but complementary sources: (i) the pathbreaking work of Joseph Schumpeter, and (ii) recent contributions to physics, chemistry and theoretical biology. It turns out that problems which are firmly rooted in the economic discipline, such as innovation, technological change, business cycles and economic development, contain many clear parallels with phenomena from the natural sciences such as the slaving principle, adiabatic elimination and self-organization. In such dynamic worlds, adjustment processes and adaptive behaviour are modelled with the aid of the mathematical theory of nonlinear dynamical systems. The dynamics is defined for a much wider set of conditions or states than simply a set of competitive equilibria. A common objective is to study and classify ways in which the qualitative properties of each system change as the parameters describing the system vary.

The origins of the present volume may be traced to comparative studies of spatial and economic dynamics which were initiated at the International Institute for Applied Systems Analysis (IIASA) in 1982. Among other tasks, the international reference group for these studies provided insights into the theoretical development of models and methods suited for the dynamic analyses of economic change processes. Many of the contributions contained herein have come from members of this reference group who have been undertaking basic research into dynamic processes such as competition, economic development and spatial adjustments.

Most of the papers were presented during four special sessions held at the 5th International Conference on Mathematical Modelling, on the campus of the University of California in Berkeley, California from July 29-31, 1985. Three additional chapters have been prepared by those authors who were invited to contribute to these sessions but could not attend. Thus the volume is as complete and unified as possible.

The editorial work has been undertaken by the Centre for Regional Science Research (CERUM) at the University of Umeå.* The work has been supported in other ways by the Department of Economics at the same university as well as by IIASA and the University of Karlstad. In particular, Jenny Wundersitz (CERUM) coordinated all editorial tasks and Ingrid Lindqvist (University of Karlstad) prepared and revised the manuscript. For their perseverence and perceptive attention to detail, the editors are sincerely grateful.

Umeå, February 1987

David Batten
John Casti
Börje Johansson

* The work on this volume has been done within a project supported by Riksbankens Jubileumsfond, grant 84/266:1.

Table of Contents

Chapter 1

Economic Dynamics, Evolution and Structural Adjustment

B.JOHANSSON, D. BATTEN and J.CASTI

From perspectives represented in this volume, contemporary economics may be regarded as a mature discipline. It has brought the state of the art close to the boundaries of the predominant paradigm. Breaking through these boundaries implies a search for alternatives or at least some alterations to this paradigm. Many of the contributions contained herein spring from such ambitions. The different chapters consist of various attempts to permeate the boundaries or to open up new pathways within the existing paradigm, and in several cases to break with the established tradition.

On occasions, economic theory and modelling has adopted concepts and research strategies from the natural sciences, in particular from classical physics. A growing circle of economic theorists have recently been inspired by some other developments within the natural sciences. In this volume one can certainly identify influences of this kind. Approaches and concepts introduced in the models are inspired by recent contributions to physics, chemistry and theoretical biology. Moreover, techniques for rendering the analysis of dynamical systems more tractable have been imported and adapted from the natural and engineering sciences. However, the problems so addressed are firmly rooted in the economic discipline with a special focus on technological change, business cycles, economic development and growth. The pathbreaking work of Schumpeter is a mutual source of inspiration for many of the authors. In this spirit the whole collection is concerned with the dynamic analysis of market economies and competition in the marketplace.

1. MODELS OF ECONOMIC DYNAMICS

The two chapters on dynamics to be found in Samuelson's "Foundations of Economic Analysis" (1946) reveal that the evolution of most theoretical ideas and principles generally undergoes long incubation periods. Samuelson's explanations and interpretations of comparative statics, stability of equilibria, cycles and parametrization of dynamic processes contain many parallels with phenomena such as the slaving principle, adiabatic elimination, and self-organization (Haken, 1978; 1983). However, his discussion concentrates primarily on two aspects of economic dynamics, namely the stability of equilibria and business cycles. As such, Samuelson's contribution has little to say about economic growth, which has basically been a post war field of interest.

Walrasian equilibrium models - in the guise of general competitive analysis - gained some momentum in the beginning of the 1950s. In retrospect one may recall a comprehensive research programme focussed on the existence of competitive equilibria, safeguarded by equilibrium prices. The associated dynamic analysis was therefore preoccupied with the stability of such equilibria. A great deal of theoretical effort was spent on the study of "artificial" price adjustment processes using the method of Lyapunov (see, e.g., Arrow and Hahn, 1971). The term "artificial" is justified here by the fact that the studies were not principally concerned with adjustment processes themselves, but rather with the stability of the essentially static notion of a competitive equilibrium. Indeed, the extent of any dynamic analysis was generally constrained by the equilibrium notion of the research programme. It may also be characterized as an underpinning for the static equilibrium analysis. As emphasized in the early work of Samuelson (1946), a static market equilibrium is of general interest only if it is stable. One might add that a competitive equilibrium will have little importance if it is not structurally stable or generic. Metaphorically speaking, if we repeat the market experiment under approximately the same conditions we would like to obtain approximately the same results (compare Hirsch and Smale, 1974).

Growth theory comprises the process of capital accumulation intertwined with increases in production and consumption. The reconstruction of the world economy after the Second World War provided stimuli for analyzing growth problems. Von Neumann's German version of the influential "A Model of General Economic

Equilibrium" (1945) initially appeared in 1937. This model and its successors (see, e.g. Georgescu-Roegen, 1951; Koopmans, 1964; Morishima, 1969) describe a multisectoral economy using linear activity analysis. The resulting equilibrium solution depicts a system in which outputs and inputs grow at the same proportional rate. In comparison with related models of multisectoral growth (e.g. Leontief et al, 1953; Johansen, 1960), the von Neumann framework represents the passage of time in the form of ageing vintages of capital goods.

During the 1950s aggregate growth models were developed along several lines. The neoclassical economic growth models, whose origins may be traced back to Ramsey (1928), characterized the optimal rate of growth and saving in an economy. The essentials of this growth theory are summarized by the following differential equation

$$f(k(t)) = c(t) + \alpha k(t) + dk/dt \qquad (1)$$

where all variables are measured per unit of labour, and where k denotes capital, f(k) output, c consumption, αk capital maintenance and dk/dt the net increase of capital (compare e.g. Solow, 1956; Phelps, 1961). One side of this growth model relates to control theory formulations. The other side takes the form of aggregate business cycle models, such as multiplier- accelerator models. Cyclical behaviour generally stems from nonlinearities affecting the investment process in an aggregate difference/differential or integral equation. From the 1930s economists experimented with this type of analysis using difference equations (see Kalecki, 1971). In a majority of cases the difficulty of finding a solution was avoided by constraining the final structure of the model to a manageable form. For example, much of the work initiated in the 1950s escaped such difficulties by retreating to the area of linear oscillators.

In the early contributions of Kalecki (1935) and Goodwin (1951) we find accelerator-multiplier models of the mixed difference-differential type, including lags and nonlinearities. The cited Goodwin model also contains a given rate of technological progress. We note in particular that the system is inherently explosive. However, structural characteristics such as nonlinearities, ceilings, and floors keep the cycle within a given range of values. The latter property, which plays a significant role in Hick's model (1950), is illustrated in Figure 1. In Section 4 we return to more recent model formulations in which nonlinearities not only replace the floor and ceiling constructions but also introduce the possibility of bifurcations and chaotic behaviour.

The contributions collected together in this volume focus on cyclic fluctuations and technological development rather than on economic growth per se. These distinctions can be readily appreciated if the notion of growth is taken to signify "more of much the same", whereas the notion of development is taken to describe a highly evolutionary process (see Batten, 1985). Indeed evolution and development are almost the same word (Boulding, 1981).

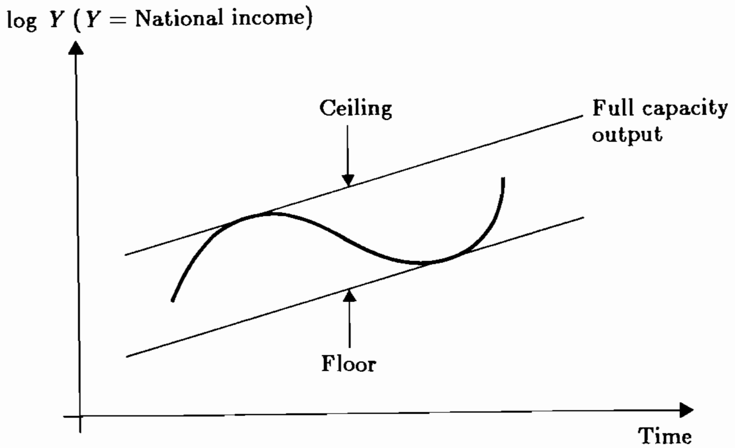


Figure 1 Illustration of floor and ceiling in Hick's model

In this introductory chapter we do not intend to examine all existing models of economic dynamics. Thus far we have mentioned some of the historical building blocks used in the world of dynamic analysis. We have chosen to introduce those particular features which are important to the collection of contributions in this volume. In subsequent sections, we shall refer to several other approaches which are also relevant. One example is the family of putty clay models in which machines of different vintages have different technical characteristics, so that old vintages gradually become economically obsolete (see, e.g., Solow, 1962; Bardhan, 1969 and 1973; Johansen, 1972). Models describing the choice of technique (Morishima, 1969), technical progress and the diffusion of technical change (e.g. Arrow, 1962; Mansfield, 1961; Hahn and Matthews, 1964; Stoneman and Ireland, 1983) are also introduced. Microeconomic aspects are represented by market-simulation models based upon recursive programming and adaptive behavior (Day and Grove, 1975). This group contains response mechanisms found in oligopoly models (Hosomatsu, 1969; Kirman, 1975) and considers the formation of expectations.

2. EVOLUTION, INNOVATION AND SYNERGISMS IN ECONOMIC DEVELOPMENT

The processes of long term change described by the classical economists such as Smith, Ricardo and Malthus appear in retrospect to rely solely on population increase and capital accumulation as their main ingredients (Baumol, 1959). By way of contrast, in the work by Marx technical change is given a more prominent place. When we read Schumpeter, evolutionary aspects of economic development are brought into the picture, led by concepts like innovation, industrial mutation and creative destruction. The association between clustering of innovations and economic cycles may be an important reason for the revived interest in Schumpeterian theories during the 1970s and 1980s.

Schumpeter's notion of innovation refers to several aspects of novelty such as the emergence of new needs and changing preferences as part of social learning processes, the development of new products satisfying established consumer needs, the use of new products and equipment in changing and improving production processes, the adoption of new organizational strategies and the opening of new markets. The foregoing collection of Schumpeterian ideas is represented in this volume by the contributions of Andersson, Day, Batten and Akin (Chapters 2-5).

Inspired by such perspectives, several scholars have promoted an approach to the modelling of economic systems under the banner of "evolutionary economics". In Nelson and Winter (1982), the term evolution refers to processes of long-term and progressive change. They also stress that the concentration on long-term and continuing elements in the economic process does not preclude that change may be very rapid. A unifying concept along these lines is that economies develop or change because they are "out of equilibrium". This idea is in keeping with the theory of self-organizing systems (Nicolis and Prigogine, 1977) and synergetics (Haken, 1983).

An essential feature of self-organizing systems is that temporal and spatial patterns evolve endogenously without being imposed on the system from outside. Of special importance is the interaction of coupled subsystems as illustrated by the following pair of differential equations

$$\dot{q} = N(q) + vK(q,z) + f(t)$$
$$\dot{z} = M(z) + vH(q,z) + g(t) \qquad (2)$$

where v may be regarded as a control parameter describing the strength of the interaction, and where f(t) and g(t) represent driving forces. Self-organization may be caused by (i) a change in the global impact of the surroundings as expressed by f(t) and g(t), (ii) an increase in the number of components as expressed by a switch from $v = 0$ to $v > 0$, or (iii) a sudden change in control parameters when the system is translated to a new state under new constraints.

Self-organizing systems are characterized by nonlinearities and their evolution involves the loss of linear stability, and the onset of transitions and chaos. Aspects of this type of behaviour are treated in this volume by Day (Chapter 3), Silverberg (Chapter 6), Puu (Chapter 10), and Dendrinos and Sonis (Chapter 15).

Synergetics has been proposed by Haken (1983) as a general theory of the dynamic behaviour of systems with particular characteristics. It deals with the cooperative interaction of many subsystems thereby engendering macroscopic systems behaviour of a self-organized nature, and therefore constitutes an important background to many of the contributions in this volume. The focus is on critical points where the system changes its macroscopic behaviour and may undergo non-equilibrium phase transitions, including oscillations, spatial structures and chaos. Among other things, this means that the sphere of interest is not merely restricted to transitions between equilibria and equilibrium-like attractors as limit cycles. The ambition is also to capture other transitions without a specific final form.

In Chapter 2, Andersson portrays creativity and the development of knowledge as a synergetic process involving a search for essential nonlinearities in the evolution of economic systems. On the macro level, synergies take the form of nonlinear interactions between sectors producing knowledge. The system described by Andersson is built up using subsystems operating on different time scales. In the literature on synergetics such dynamic processes are treated by means of the slaving principle.

Consider the following prototype system in which the y-variable adjusts more quickly than the x-variable:

$$\dot{x} = ax - xy$$
$$\dot{y} = -by + x^2. \qquad (3)$$

When the coefficient a is positive and small, and $b \gg a$, we may use the following approximation: $dy/dt \approx 0$ which yields $y = x^2/b$ so that y is "slaved" by x. This procedure represents one way of reducing the dimension of a complex system in order to investigate its qualitative behaviour at critical points. In Andersson's case, "creative explosions" may occur at such points.

In Chapter 3, Day presents a series of arguments which are strongly related to synergetics. In his panorama the economy is a complex adapting system, adjusting in disequilibrium on the basis of obedience, imitation, formation of habits, and experimentation. Aggregate patterns which are brought into focus include waves of productivity change and fluctuations in output, a succession of epochs of economic organization and switches between technological regimes.

As the name suggests, there is a nexus between evolutionary economics and biology as regards the genetic process of variation interacting with forces of individual behaviour and environmental selection. In his outline of (M,R)-systems (Chapter 13), Casti is explicit about such a heritage when suggesting the metaphor of a global industry as a living multicelled organism. In his conceptual framework, the concepts of metabolism, repair and replication are the building blocks in an analysis emphasizing the functional rather than structural organization of an economy. Akin makes a similar statement in Chapter 5 about biology affinity. He investigates and develops a model due to Nelson and Winter (1982); it is a model of competitive growth among firms in an industry and is claimed to be inspired by evolutionary biology. It is not surprising then that mathematical techniques developed originally in population genetics can be applied to solve the model of competition formulated by Akin.

In Batten's fable for growth merchants (Chapter 4) many of the ideas appearing in other chapters are interwoven. The story is formulated around the paradigm of self-organization, thereby implying that the dynamics of qualitative change should be endogenously determined. Following Schumpeterian lines, it is asserted that the process of industrial mutation revolutionizes the economic structure from within and can usefully be interpreted as a process of creative destruction. In Table 1, some of the major components of the fable are summarized.

Table 1 Facets of economic development (see Chapter 4)

Catalyst of industrial change	Knowledge creation and diffusion, R&D
Generation of new alternatives	Development of mutants or innovations in the form of new production processes (techniques) and new products
Exit and entry triggers	Existing production units and products exit from the market as profits decline; entry is stimulated by profit opportunities
Selection mechanisms	Evolution by competitive substitution. The best practice techniques which are superior at a given point in time are multiplied through imitation and inferior techniques are eventually updated or phased out. New products with a different attributes-price combination replace old ones gradually as they are preferred by customers and the output is increased through investments in production capacity
Time paths	Within the constraints imposed by the evolving size of each market, every successful new variant penetrates its chosen market and follows a logistic time path, thereby creating successive waves of innovation – imitation each time superior variants (techniques and products) are introduced
Structural patterns	At each point in time, techniques and products of different rent vintage will coexist and form a particular distribution of productivity, profit and capacity levels over firms or production units. Each such distribution will reflect in a transparent way the historical evolution and the conditions for furture change in each market.

3. VINTAGES, SUBSTITUTION AND STRUCTURAL ADJUSTMENT

Table 1 contains a description of structural patterns. Such a pattern may reflect the distribution of technical vintages of machines or plants in an industrial sector. It may also reflect the distribution of productivity differences between such vintages. Such a structural pattern can also describe how the share of total output or employment in the economy is divided between sectors. For a given type of structural pattern, we may define sudden or gradual changes to the pattern as structural adjustment. The term adjustment indicates that the simultaneous change of several variables or subsystems arises out of interdependent responses or mutual adaptation.

In economics and related social sciences, structural patterns have been associated with variables changing at a slow pace from a systems perspective. This means that structure is a concept which depends on the aggregation level and the subsystem under study. In particular, the study of structures has been occupied with the examination of a set of variables x_i and $y_i = x_i/\Sigma x_i$, $i \in I$ for which, during given time periods,

$$\dot{x}_i \neq 0$$

$$|\dot{y}_i| \leqq \varepsilon \qquad (4)$$

where ε is a small number. Hence, the research has focussed on system properties which remain approximately invariant over time intervals which are long relative to the chosen time scale. In many cases the variables y_i characterize a system and slave other variables of the system. It is then self-evident that such a modelling strategy must be complemented by an examination of how sudden changes in the y_i-variables may be triggered.

Patterns of the type discussed here may refer to the distribution of economic activities over locations or the spatial distribution of the price of an economic resource, as discussed by Beckmann in Chapter 14. Alternatively, they may refer to the distribution of profitability levels accruing to production units in an industry as described by Johansson in Chapter 7. In the modelling of economic growth, there has been a tradition of allocating substantial effort to "simple" structures like K/Q and K/L, where for a sector or the complete economy K denotes capital, L labour, and Q output. Special interest has been shown in technical change which leaves one or several of these ratios unchanged over time. Such invariances are said to be neutral vis-à-vis the technical change (compare, e.g., Saito, 1981).

Several of the chapters in this volume analyse structural adjustments in the composition of product vintages and vintages of production techniques. The concept of a vintage is (from certain perspectives) one of the most fundamental dynamic concepts to have emerged from the body of economic thinking. Its potential importance relates to the fact that it reflects a lag or delay structure and permits a direct comparison of the relative speed of change and degree of obsolescence for quite diverse variables pertaining to an economy. In addition, there are rich opportunities to relate vintage information to empirically observable indicators.

To this point, the vintage concept has played a relatively minor role in economic research. In the early 1960s a number of economists began to formulate growth models in which capital equipment was characterized by its date of construction, in the sense that the equipment embodies the best available technology (best practice technique) of its date of construction. This implies that factor combinations (inputs to the given production process) are variable before the investment takes place (ex ante) and fixed once capital has actually been committed (ex post). This type of analysis was pioneered by researchers such as Svennilson (1944), Johansen (1959), Salter (1960), Solow (1960) and Phelps (1962).

In the above approach a sector may be portrayed as a distribution of production capacities, each with a specific production technique reflecting its vintage. This is illustrated schematically in Figure 2 where each vintage is represented by a vertical line. The symbol x measures capacity, and a and b represent the amount of two different inputs per unit capacity. Similarly, we may interpret the vertical lines as sales volumes for different products belonging to a product group. In this case, the axes a and b are used to measure the amount of two selected attributes per unit of each product. These commodities will in general correspond to different vintages, some commanding an increasing share of their market and others gradually losing their market share.

If we define a sector as a set of establishments or production units, we may then recognize that each unit consists of plant and equipment of different durability and with different susceptibilities for replacement by a newer piece of equipment (compare a factory building with a truck). Other changes in the organization may not require extensive replacement of the capital structure. It is also true that every production unit may change its product mix (goods and services) by modifying the bundle of attributes of each product.

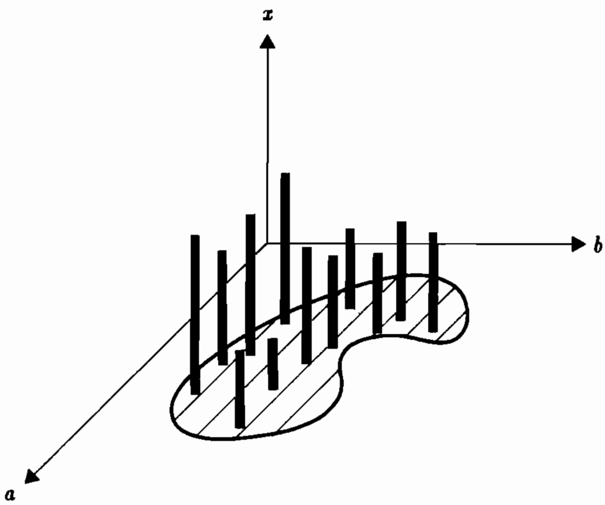


Figure 2 Illustration of discrete vintage distributions

All the above changes alter the vintage structure of the production unit, and all require resource commitments in the form of investments. Having reached these conclusions, two important aspects of the associated dynamics may be noted. First, a distinction may be made between physical depreciation and economic obsolescence. The latter is an empirical observation in the form of a successively decreasing stream of gross profits (quasirents). Competition from new plants or renewed plants using more economically efficient plant or equipment is one of the main reasons for the decline in the gross profits of older plants. New vintages take the form of new production techniques and new products. Once an investment in a new vintage has been made, it represents a sunk cost. Hence its ability to compete is constrained only by current variable costs. On the other hand, the decision to introduce a new vintage must be based on a calculation including both fixed and variable costs. In other words, there is a significant asymmetry between the conditions for exit and entry of vintages.

Our sketch of vintage renewal corresponds to the repair mechanism in Casti's metabolism – repair system in Chapter 13. A production process of a given vintage is identified as a metabolic "machine" composed of a production and a sales map. In this system, the capacity to repair and innovate is represented by a repair map, embodying the "genetic" capability to renew the metabolic process. In Chapter 6 Silverberg presents a simulation model of firm competitiveness based on embodied technical progress in which the vintage replacement is not constrained to steady-state development paths. In this world the average productivity of a sector is affected both by scrapping old vintages and investing in new ones. During downswings in a business cycle these two phenomena will be simultaneous, while upswings mainly affect capacity increase. This is also emphasized in Chapter 7, where Johansson describes the interaction between exit and entry of capacities of different vintages on the one hand, and the development of demand and price formation on the other.

In Schumpeter's "Theory of Economic Development" (1934), a market with few and modest innovations is assumed to converge towards perfect equilibrium, which may be disrupted again as new and radical innovations enter. In this analysis, the entrepreneurs are motivated by the possibility of gaining a temporary monopoly by exploiting their innovation before their competitors. If successful, they receive a premium for being first. The size of this extra profit depends on the length of the delays in the response from competitors. These responses have the form of imitation processes and the development of alternative innovations. It has been argued by Day (1982) that the market only provides an innovation-promoting milieu if it persistently stays out of perfect equilibrium. Although the model formulations vary, in Chapters 4-7 we can recognize different versions of this idea. In particular, innovations and imitations are assumed to reflect entrepreneurs' attempts to obtain temporary monopoly-like profits. This image of markets adapting to "disequilibria" is clearly related to the idea of synergetic, self-organizing systems.

Vintage and innovation theories have close linkes with product cycle theory as presented by Dean (1950), Vernon (1966), Hirsch (1967) and Pasinetti (1981). This theory attempts to explain both product development and process improvements. In the perpetual race for temporary monopoly profits and attempts to safeguard achieved market positions, standardization of both products and production techniques gradually tends to change the input structure (or production function) of a given type of commodity in a predictable way. In this type of model, the imitation and exit of obsolete vintages gradually forces an expanding product innovation away from an initial situation of dynamic competition towards price or cost competition of the classical type. However, the introduction of new products (which may constitute equipment for other processes) gives birth to further evolution. This phenomenon can be modelled as a succession of substitution processes (Batten and Johansson, 1985) in which new and modified products, services, sales and distribution systems replace old ones. Some basic properties of substitution and diffusion processes of this kind are examined by Sonis in Chapter 8.

Economic evolution in market economies brings several types of cyclical elements into focus. One such phenomenon is repetitive switches between "dynamic competition" and "price competition". These dynamic patterns may be associated with business cycle behaviour. An example is provided in Chapter 9 by Haag, Weidlich and Mensch. The innovation and diffusion processes also possess fundamental spatial dimensions as described by Blommestein and Nijkamp in Chapter 16.

4. PERIODIC AND CHAOTIC BEHAVIOUR

This introductory chapter aims to provide a conceptual framework for the contributions in this volume. In the preceding sections we have paid attention to nonlinearities, delay structures, instabilities and synergetic properties in general. Among economists and other social scientists, the pertinent problems have been investigated using dynamic models operating on both discrete and continuous time-scales. Considerable attention has been paid to the fact that discrete-time models move quite easily into a domain of chaotic behaviour as illustrated in May (1974 and 1976) and Day (1981). Just as with the continuous versions, the discrete-time models are sensitive to the size of reaction parameters ("speed" of reaction). In addition, these models are sensitive to the choice of periods and the chosen lag structure.

Two important conclusions can be drawn from this. First, meaningful discrete models must be formulated on the basis of a careful selection of lag structures to accurately represent the studied system. Otherwise they may too easily generate chaotic behaviour which is not representative. Second, modellers and decision makers in the economic system observe "reality" by means of periodically gathered statistics of various forms. This latter observation may be important when formulating models in continuous time. A related conclusion which has been stressed by Day is that when macroeconomic information is only available over discrete time periods, and the estimated model displays turbulence, then we cannot rely on econometric methods to make predictions. This may be a useful starting-point for studies of "regularities" in systems displaying chaotic behaviour.

The classical business cycle models were formulated in a discrete time setting with the help of a multiplier-accelerator principle. Here the multiplier component describes how consumption is stimulated by the level of national income one period back, and the accelerator component describes current induced investment as a response to the change in consumption between the two most recent periods. This type of model was examined by Samuelson in 1939. The stability and cyclic behaviour of this family of models depends critically on the relative size between the two parameters describing the multiplier and accelerator reactions. In an attempt to develop a more realistic model, Hicks (1950) introduced more elaborate lags and related investment to total demand rather than to consumption. He also added an autonomous growth factor. Evaluating the parameters of his model against observed statistics, Hicks was obliged to conclude that his model was explosive. For this reason, he introduced a ceiling and a floor as depicted previously in Figure 1. These devices may be thought of as non-linear elements establishing an upper and lower limit to income and generating cycles of constant amplitudes around the growth term (in relative terms).

An approach related to Hicks' contribution is found in Goodwin (1967). In this case a continuous time setting is adopted. The nonlinearities corresponding to Hicks' ceiling and floor are introduced by means of Lotka-Volterra equations which had been developed earlier to describe the interaction between a predator and its prey. In a metaphoric way, the two species appear in Goodwin's model as the mutual influence of wages and investments generating a cyclic pattern. This heritage, in particular, is represented in Chapter 11, where Glombowski and Krüger elaborate on the original Goodwin model. In Chapter 10, Puu develops the treatment of nonlinearities further in a multiplier-accelerator framework. He also provides examples showing how complex dynamics and chaotic patterns may emerge.

We have argued from the outset of this chapter that for a long period economic research has been under the lock and key of a competitive equilibrium notion that has generally obstructed dynamic analyses. A recent study which strives to shed light on cyclic behaviour within the competitive equilibrium framework may be found in Grandmont (1985). He describes a discrete-time market economy which, in a stationary environment, can generate a sequence of competitive equilibria forming a cycle (periodic orbit). This pattern is endogenous and arises from the conflict between a wealth effect and an intertemporal substitution effect in an economy with overlapping generations. Cycles of different periods will typically coexist and the occurrence of these cycles is the result of a bifurcation-like phenomenon. The aim of the paper is to illustrate that cycles can persist in an economy satisfying the conditions of competitive equilibrium, and that such cycles must not and need not be imposed from outside.

Scientific disciplines are indeed evolving systems for which synergetic properties may be of special importance. From this perspective, the contribution by Grandmont is illustrative and complements some of the work in this volume. It also draws our attention to the potential rewards which may accrue from an association between micro and macrooriented models of systematic fluctuations. Such a strategy is clearly related to Haken's suggested approach. Although the starting-point differs, one can recognize elements of this framework in Chapter 9 (by Haag, Weidlich and Mensch). Their model operates with a stylized micro- specification and generates aggregate macropatterns.

The above model is said to be based on Schumpeterian concepts, and thus features innovators and imitators as the prime movers, creating microeconomic differences among producers. At any point in time this corresponds to a picture of monopolistic competition of the kind found in Chamberlin (1938). In the model, investors have propensities to allocate their investment capital to expansionary, E-type, and rationalizing, R-type, investments. Microshifts in individual propensities combine to generate swings in the aggregate economy reflected by the aggregate investment structure index $Z(t) = (E(t) - R(t))/(E(t)) + R(t))$. The transitions of investors between the two types of investments are activated by two parameters, one called the 'alternator' and the other the 'coordinator'. A dynamic specification of the first parameter is necessary to keep the Schumpeter clock ticking over forever. The final choice of such a specification is left open. The most noteworthy feature of the model construction is the linking between microevents and macropatterns, which is obtained by means of a master equation formulation of the microsystem.

In Chapter 10 by Puu, basic knowledge about the economic system is also used to establish constraints or principles for the behaviour of the economic system at the aggregate level. Applying strategies of this kind enables us to make use of established economic theory when introducing new mathematical techniques in the analysis of the dynamic behaviour of economic systems. The heritage from economic theory represents accumulated knowledge and experiences over several centuries and has been sadly disregarded in many recent contributions to economic dynamics.

The background to the Puu model is aggregate multiplier-accelerator analysis which was discussed earlier in connection with the contributions by Hicks and Goodwin. As observed then, linearization of the pertinent differential equations brings about explosive cycles and exponential growth. The introduction of a nonlinear investment function prohibits such solutions. The investment function relates the investment level to changes in income (or demand), dY/dt. In this case, nonlinearity means that as dY/dt becomes very large the increase in the investment level becomes asymtotically zero. With this assumption (which may be thought of as a system consistency property) the existence of a limit cycle is established. At this point two interacting economies are studied. Both behave in accordance with the multiplier-accelerator principle. One of the economies is assumed to be large enough to drive the smaller one. As a consequence the smaller economy is represented by a forced nonlinear oscillator. In general the interacting frequencies will be incongruent. The outcome is various complex combinations of cycles and, in particular, chaotic motion.

The model by Glombowski and Krüger (Chapter 11) contains a series of exercises with an enlarged version of Goodwin's predator-prey model. They introduce a general model with a large number of aggregate economic relations establishing conditions for the dynamic behaviour of the model. By means of a systematic variation of the coefficients of the general model they obtain different variants with a reduced number of equations. Each case is assessed with regard to equilibrium values of steady state solutions and cyclic motions. The numerical experiments not only simulate closed orbits but also cases with slowly explosive cycles. Among the cases studied the authors manage to formulate variants which display frequencies and amplitudes (cycle periods) which conform more accurately to empirically observed cyclic patterns than the cycles which are generated by the original Goodwin model. The variations in the model concern the endogeneity and neutrality of productivity change, technical progress, and so forth. Although the authors do not use such arguments, one might interpret their various reduced model variants as being generated by assumptions that the eliminated parts of the system equilibrate rapidly relative to other parts. Moreover, as the general approach has some affinity with "experimental mathematics" (Chapter 15), it demonstrates the difficulties in summarizing the results and interpretations in a "tractable" or conclusive form.

In Chapter 12, written by Medio, we return to a multiplier-accelerator setting displaying certain similarities with the approach in Chapter 10. Medio introduces a multisector model $x = Ax+Bx$ where x is an activity vector, A an intermediate delivery matrix and B a capital increment matrix. This model is modified to allow for discrepancies between actual and desired levels. The equilibrium conditions are replaced by an adjustment mechanism with a lag structure. In a fashion similar to Puu, Medio formulates a nonlinear investment response. The stability of the system is studied and conditions for periodic solutions are established by means of Lyapunov functions and bifurcation theory. The results turn out as follows: For a sufficiently weak accelerator, stability obtains. Moreover, there exists a critical value for the accelerator; in this neighbourhood a Hopf bifurcation will occur and a family of periodic solutions will exist. The nature of the cyclical oscillations is not examined but they are illustrated numerically.

The presentation by Puu illustrates that the interaction between economies may have its own impacts on economic fluctuations (see also Puu, 1986). Beckmann's contribution (Chapter 14) presents examples of economic processes in space which generate wave and diffusion equations. A key point here is the importance of how space is represented. Two variants are considered: (i) explicit distances and (ii) local interaction implying that interdependencies fade away quickly as distance increases. Various aspects of spatial dynamics are also treated in Chapter 16 by Blommestein and Nijkamp. In their case the change processes are generated by the adoption and diffusion of innovations.

Chapter 15 by Dendrinos and Sonis concentrates on relative dynamics in discrete time with reference to spatial processes such as the dynamics of population stocks over spatial locations when driven by comparative advantages. They discuss examples which conform to the slaving principle. One such case describes fast population dynamics dictated by the current values of slowly changing parameters representing location advantages. They also mention the possibility of representing the speed of change in stocks by means of lag structures. With its focus on turbulence in location dynamics, their paper constrasts the processes of diffusion and competitive exclusion outlined by Sonis in Chapter 8.

5. MARKET COMPETITION VIEWED AS A DYNAMIC PROCESS

A superficial glance at the contents list of this volume might cause the reader to conclude that the volume consists of a number of disconnected ideas and models. In the preceding four sections we have endeavoured to outline some basic concepts and perspectives which hopefully identify many coherent elements and general principles which pervade all these contributions and point to future work. One unifying observation is that all the chapters deal with dynamic processes in market economies. Incentives to invest and to adjust prices, production volumes and consumption are clearly related to our perception of market economies. One group of chapters focusses specifically on the competition between firms, while another group examines aggregate models which are themselves based on assumptions about such competition. Associated with this one finds an even stronger focus on changes in production structure and in the composition of techniques and products. The aggregate models refer to growth and cycles of the production volume or economic activity.

In section 1 we identified competitive equilibrium analysis as a mainstream research programme which has persisted for several decades since its inception around 1950. Such a widespread research programme corresponds to the vigorous stage of what Kuhn (1970) has described as a paradigm. Unfortunately, a paradigm may also achieve a state of maturity in which 'research' no longer dominates, and conventional wisdom is preserved via ceremony, defence of the established theoretical body and hostility against heretics. With its focus on competition in a market system, the above-mentioned general equilibrium paradigm may more specifically be referred to as Walrasian or neo-Walrasian market theory. In this tradition the analysis is more emphatically nondynamic than in the parallel tradition of Marshall or in the thinking of Adam Smith.

In recent decades we can identify many economists who have tried to modify or break away from the Walrasian tradition. Frequently they have related their efforts to the work of scholars such as Smith, Marshall, Chamberlin, Keynes and Kalecki (see, e.g., Negishi, 1985). In particular, Schumpeter constitutes a prominent pillar and popular source for such work. This is also the case for this volume. However, the reader may also discern a similar kind of reference to certain (incomplete) aspects of model formulations by Hicks and Samuelson.

This volume provides some small samples and suggestions of new directions in economic modelling. The joint concern is with the type of dynamics considered, the pertinent dynamical formulations and the appropriate techniques for analyzing and solving the models. One characteristic which is recurrent in the various chapters relates to those nonlinearities which reflect the basic properties of a changing market economy. An associated characteristic is the desire to depict the market as a dynamic process in which one can identify competitive phenomena such as survival of the fittest, an ongoing search for market niches, and the development of specialization. Other elements of common interest include creativity, knowledge creation, innovations and imitations as well as the occurrence of production in time patterns that resemble the product cycle. Development and maturity yielding routinization, market extension and increasing scale of production are also of mutual concern.

In this image of the market system, the emergence of novelties is combined with the adoption of decision rules (including production techniques) which gradually become rigid routines and are eventually replaced by new ones. In such a system, adjustment processes and adaptive behaviour are important elements of the dynamics. Such processes of synergetic interaction lead to self-organizing behaviour. From such a novel perspective, the contributions in this volume still appear primitive. We still find very few attempts to introduce systematic aggregation procedures. The ensemble of chapters illustrate a clear need to develop frameworks in which properties of macrosystem dynamics can be imposed effectively as constraints and conservation principles for microsystems, and in which properties of disaggregate, subsystem or microsystem dynamics can be used in the derivation and formulation of aggregate dynamic models.

In several of the chapters discussions about competitive equilibrium notions of the Walrasian type are included. It has been argued that evolutionary economics and models of structural adjustment must pertain to disequilibrium processes. In our view, such claims may be superfluous. By the way dynamics is portrayed in this volume, response and adjustment patterns are underpinned by specific arguments and assumptions. These assumptions may indeed be derived from standard economic theory and conventional thinking. For example, the composition of production factors may adjust to their respective marginal productivities (compare Puu, 1986). Often the resulting processes do not generate paths along which the conditions for a competitive equilibrium are satisfied. The possibility of such trajectories is a natural consequence of choosing formulations with dynamic equations. First, the dynamics is defined for a much richer set of states than just the set of competitive equilibria. Second, necessary system consistency conditions frequently imply that the model develops patterns which may be characterized as processes of economic disequilibrium. The point we are stressing here is that the basic assumptions should mainly concern the dynamic behaviour rather than the structures which follow from the dynamic processes.

One way of establishing a bridge between the field of 'evolutionary economics' and more orthodox economic theory would be to suggest that innovations bring about new products and production processes which mature along a development path resembling a product cycle. Near the outset of this path, dynamic competition of the Schumpeterian type is the dominating characteristic. During the second phase the production evolves in the direction of price competition in which standardized product attributes and routinized behaviour reflect the type of price-taking behaviour which corresponds closely to orthodox theory. We may illustrate this suggestion with the help of the following table which compares some conjectured properties of dynamic and price competition. It is hoped that the set of chapters built around this theme may stimulate further innovative thinking in the Schumpeterian spirit.

Table 2 Price and dynamic competition

PRICE COMPETITION	DYNAMIC COMPETITION
Price sensitivity in product markets is high; discounting may prevail	Price sensitivity in product markets is low; price-making strategies prevail
Market forms resemble sales by auction; markets are global and saturated	Sales by orders; customer-adapted products; markets are segmented and immature
Large-scale production techniques, standardized competence and routinized jobs	Adaptive production techniques where knowledge-orientation, creativity, and originality dominates
Standardized techniques adjust to export cycles and a race to minimize production and delivery costs	Innovations and imitations occur in response to import cycles and novelties in the market place
R&D is low and oriented towards process improvements and marketing	R&D is high and oriented towards product developments and refinement
Productivity is determined by cost efficiency, and adjustment of capital	Technical and design advantages as well as patents generate temporary monopoly-

equipment (capital intensity)	like profits on the basis of nonmaterial investments

Adapted from Johansson and Strömqvist (1986)

REFERENCES

Arrow, K.J., The Economic Implications of Learning by Doing, Review of Economic Studies 29, 1962, pp. 155-173.

Arrow, K.J. and Hahn, F.H., General Competitive Analysis, Holden-Day, San Francisco, 1971.

Bardhan, P., Equilibrium Growth in a Model with Economic Obsolescence of Machines, Quarterly Journal of Economics, 83, 1969, pp. 312-323.

Bardhan, P., More on Putty-Clay, International Economic Review 14, 1973, pp. 211-221.

Batten, D.F., Introduction, in Batten, D.F. and Lesse, P., (eds.), New Mathematical Advances in Economic Dynamics, Croom Helm, London, 1985, pp. 1-12.

Batten, D.F. and Johansson, B., Industrial Dynamics of the Building Sector: Product Cycles, Substitution and Trade Specialization, in Snickars, F., Johansson, B. and Lakshmanan, T.R., (eds.), Economic Faces of the Building Sector, Document D20: 1985, Swedish Council for Building Research, Stockholm, 1985.

Baumol, W.J. , Economic Dynamics. An introduction. Macmillan, New York, 1959.

Boulding, K., Evolutionary Economics, Sage Publications, Beverly-Hills, 1981.

Chamberlin, E.H., The Theory of Monopolistic Competition, Harvard University Press, Cambridge, Mass., 1938.

Day, R.H. and Grove, T., Adaptive Economic Models, Academic Press, New York, 1975.

Day, R.H., Complex Dynamics in Recursive Economizing Models, Transactions of the IEEE, 1981, pp. 1381-1392.

Day, R.H., Dynamical Systems Theory and Complicated Economic Behavior, MRG working Paper 215, Department of Economics, University of Southern California, 1982.

Dean, J., Pricing Policies for New Products, Harvard Business Review, 28, 1950, pp. 48-53.

Georgescu-Roegen, N., The Aggregate Linear Production Function and its Applications to von Neumann Economic Model, in Koopmans, T.C., (ed.), Activity Analysis of Production and Allocation, Chapter 4, Wiley & Sons, New York, 1951.

Goodwin, R.M., The Non-Linear Accelerator and the Persistence of Business Cycles, Econometrica 19, 1951, pp. 1-17.

Goodwin, R.M., A Growth Cycle, in Feinstein, C.H., (ed.), Capitalism and Economic Growth, Cambridge University Press, 1967, pp. 54-8..

Grandmont, J.M., On Endogenous Competitive Business Cycles, Econometrica 53, No. 5, 1985, pp. 995-1045.

Hahn, F.H. and Matthews, R.C.O., The Theory of Economic Growth: a survey, Economic Journal 74, 1964, pp. 825-832.

Haken, H., Synergetics. An Introduction, Springer Verlag, Berlin, 1978.

Haken, H., Advanced Synergetics, Springer Verlag, Berlin, 1983.

Hicks, J.R., A Contribution to the Theory of the Trade Cycle, Oxford University Press, London, 1950.

Hirsch, S., Location of Industry and International Competitiveness, Oxford University Press, London, 1967.

Hirsch, M. and Smale, S., Differential Equations, Dynamical Systems and Linear Algebra, Academc Press, New York, 1974.

Hosomatsu, Y., A Note on the Stability Conditions in Cournot's Dynamic Market Solution when neither the actual Market Demand Function nor the Production Levels of Rivals are Known, Review of Economic Studies 36, 1969, pp. 117-122.
Johansen, L., Substitution versus Fixed Production Coefficients in the Theory of Economic Growth: A Synthesis, Econometrica 27, 1959, pp. 157-177.
Johansen, L., A Multi-Sectoral Study of Economic Growth, North-Holland, Amsterdam, 1960.
Johansen, L., Production Functions, North-Holland, Amsterdam, 1972.
Johansson, B. and Strömqvist, U., Diffusion of Technology and Import Substitution, Report 1986 No 7:2, Stockholm County Council, Stockholm, (in Swedish), 1986.
Kalecki, M., A Macro-dynamic Theory of Business Cycles, Econometrica 3, 1935, pp. 327-344.
Kalecki, M., Selected Essays on the Dynamics of the Capitalist Economy, Cambridge University Press, Cambridge, 1971.
Kirman, A.P., Learning by Firms about Demand Conditions, in Day, R. and Groves, T., (eds.), Adaptive Economic Models, Academic Press, New York, 1975.
Koopmans, T.C., Economic Growth at a Maximal Rate, Quarterly Journal of Economics, 78, 1964, pp. 355-394.
Kuhn, T.S., The Structure of Scientific Revolutions, University of Chicago Press, Chicago, 1970.
Leontief, W.W. et.al., Domestic Production and Foreign Trade: The American Capital Position Re-examined, Proceedings of the American Philosophical Society, Sept. 1953, pp. 332-349.
May, R.M., Biological Population with Non-overlapping Generations, Science, No. 186, 1974, pp. 645-647.
May, R.M., Simple Mathematical Models with very Complicated Dynamics, Nature, 26, 1976, pp. 459-467.
Mansfield, E., Technical Change and the Rate of Imitation, Econometrica, 29, 1961, pp. 741-766.
Morishima, M., Theory of Economic Growth, Oxford University Press, 1969.
Negishi, T., Economic Theories in a non-Walrasian Tradition, Cambridge University Press, Cambridge, 1985.
Nelson, R.R. and Winter, S.G., An Evolutionary Theory of Economic Change, Harvard University Press, Cambridge, Mass., 1982.
Nicolis, G. and Prigogine, I., Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations, John Wiley & Sons, New York, 1977.
Pasinetti, L., Structural Change and Economic Growth, Cambridge University Press, Cambridge, 1981.
Phelps, E.S., The Golden Rule of Accumulation: a Fable for Growthmen, American Economic Review, 51, 1961, pp. 638-643.
Phelphs, E.S., The New View of Investment: a Neoclassical Analysis, Quarterly Journal of Economics, 83, 1962, pp. 548-567.
Puu, T., On Spatiotemporal Dynamics of Capital and Labour, Working paper from CERUM 1986:16, Centre for Regional Science, Umeå university, 1986.
Ramsey, F., A Mathematical Theory of Saving, Economic Journal 38, 1928, pp. 543-559.
Saito, R., Theory of Technical Change and Economic Invariance, Academic Press, New York, 1981.
Salter, W.E.G., Productivity and Technical Change, Cambridge University Press, London, 1960.
Samuelson, P.A., Interactions between the Multiplier Analysis and the Principle of Acceleration, Review of Economic Statistics 21, 1939, pp. 75-78.
Samuelson, P.A., Foundations of Economic Analysis, Harvard University Press, Cambridge, Mass., 1946.

Schumpeter, J.A., Theory of Economic Development, Oxford University Press, New York, 1934.
Solow, R.M., A Contribution to the Theory of Economic Growth, Quarterly Journal of Economics, 70, 1956, pp. 65-94.
Solow, R.M., Investment and Technical Progress, in Arrow, K.J., Karlin, S. and Suppes, P., (eds.), Mathematical Methods in the Social Sciences, Stanford University Press, Stanford, 1960.
Solow, R.M., Substitution and Fixed Proportions in the Theory of Capital, Review of Economic Studies 29, 1962, pp. 207-218.
Stoneman, P. and Ireland, N., The Role of Supply Factors in the Diffusion of New Process Technology, Economic Journal, Supplement, March, 1983, pp. 65-77.
Svennisson, I., The Growing Return to Industrial Labour in View of Swedish Experiences, Studies in Economy and History in Honour of Eli, F., Hekscher, Almqvist & Wicksell, Uppsala, pp. 238-252 (in Swedish), 1944.
Vernon, R., International Investment and International Trade in the Product Cycle, Quarterly Journal of Economics, 90, 1966, pp. 190-207.
Von Neumann, J., A Model of General Economic Equilibrium, Review of Economic Studies, 13, 1945, pp. 1-9.

PART A: DYNAMIC COMPETITION AND ECONOMIC EVOLUTION

Chapter 2

Creativity and Economic Dynamics Modelling

Å.E. ANDERSSON

NEW KNOWLEDGE IN NEO-CLASSICAL ECONOMICS

The neo-classical theory of economic dynamics focusses on the consequences of growth of labor and capital on the total level of production.

As a particularly simple example, we can use the following model from early neo-classical growth theory (Solow, 1956, Swan, 1956):

$$\dot{C} = \sigma F(C,L) - \delta C \tag{1}$$

where $\dot{C}$ ≐ net growth of capital
C ≐ stock of capital (giving a proportional flow of capital services)
L ≐ stock of employed labor (giving a proportional flow of labor services)
σ ≐ rate of investment out of production
δ ≐ rate of depreciation of capital

F is assumed to be a linearly homogenous and r-differentiable production function with $r > 2$.

Observing that $(d(C/L)/dt)/(C/L) \equiv dC/C - dL/L$

$$\dot{c}/c = \sigma (F(C,L))/C - \delta - \eta$$

where $c \equiv C/L$; $\dot{c} \doteq (d(C/L))/dt$ and $\eta \doteq (dL/dt)/L$.

Assuming $L(t) = L(o)e^{\eta t}$ we have

$$\dot{c}/c = \sigma\ F(C, L(o)e^{\eta t})/C - \delta - \eta.$$

At an equilibrium $\dot{c}/c = 0$ and hence

$$\sigma/(\delta + \eta) = \text{capital/output}.$$

Thus, an increase in the rate of depreciation or of labor should decrease the equilibrium capital-output-ratio.

To illustrate this assume F to be of the form $F = L^{\alpha} C^{1-\alpha}$. This implies that

$$\sigma/(\delta + \eta) = C/(L^{\alpha}C^{1-\alpha})$$
$$\sigma/(\delta + \eta) = C^{1-(1-\alpha)}/L^{\alpha}$$
$$\text{i.e. } c = [\sigma/(\delta + \eta)]^{(1/\alpha)}.$$

The equilibrium capital intensity (c) is monotonically increasing with the propensity to invest (σ) and decreasing with the rate of depreciation (δ) and of the growth of stock of labor (η). Production, stock of capital and labor will all grow at the same net rate η at the equilibrium, i.e. the exogenously determined rate of growth of the stock of labor.

Technological development is mostly introduced into this model in an equally simplistic fashion as a "labor augmenting factor", λ. Thus, in efficiency units, labor would grow at the rate $(\eta + \lambda)$. With the assumed function, the equilibrium rate of growth of capital and production would now be $(\eta + \lambda)$, keeping the capital per efficiency unit of labor constant at the level $[\sigma/(\eta + \lambda + \delta)]^{1/(1-\alpha)}$. The propensity to invest σ, can be made a function of the marginal productivity of capital, without any basic change to the conclusions.

In this growth model and all similar simplistic models, the focus is on the potential capacity of an economy to adapt to exogenously determined technological changes.

Technological change is thus not an endogenous economic variable in these models.

TECHNOLOGY AND ECONOMIC INTERDEPENDENCIES

The same allegation can be asserted against dynamic input-output models as well. To show this, we assume a model of an economy with n sectors in static (a_{ij}) and dynamic (b_{ij}) interdependencies:

$$x \geqq A_k x + B_k \dot{x} \tag{3}$$

where x $\doteq$ n-vector of production levels

A_k $\doteq$ n-by-n-matrix of input-output parameters at technology k

B_k $\doteq$ n-by-n-matrix of investment growth-of-output parameters at technology k.

This model can be transformed into the slightly different version (4).

$$\lambda_k x = A_k x + B_k \bar{g}_k x \tag{4}$$

where $\bar{g}$ $\doteq$ planned rate of growth

λ $\doteq$ unknown rate of capacity utilization.

A long term growth equilibrium is defined to be a state in which the planned rate of growth is such that $\lambda = 1$ and $x \geqq 0$. Let us assume that a number of different technologies are available. For each technology k, λ_k is a function of g_k as illustrated in Figure 1.

Using the Perron-Frobenius theorem for indecomposable, square, non-negative matrices, we are assured that λ_k increases with g_k for any given technology k. There is thus a maximum g*, and an associated x*, compatible with a general equilibrium for any technology k. In Figure 1, g^*_{II} is larger than g^*_I. Hence, technology II will be chosen in favor of technology I, if the equilibrium long term growth rate is to be maximized.

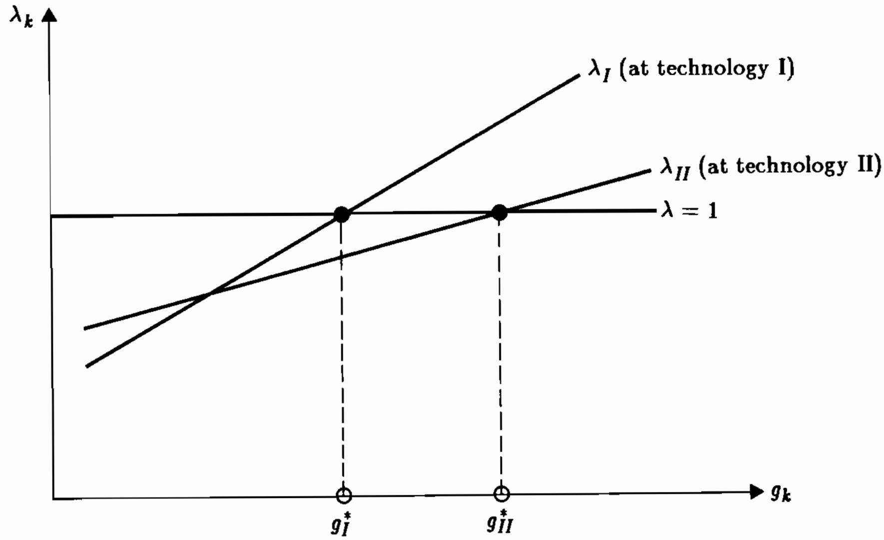


Figure 1 Long Term Equilibrium and Capacity Use with Two Technologies

In this model class, technologies are given exogenously and chosen by the economic actors so as to ensure a maximum long term equilibrium growth rate.

We can thus conclude that technological development is exogenous, both to neo-classical and inter-industrial growth models. Subsequently, the creation of new technologies will be analyzed at the macro and micro level, and basic components of models of interdependent technological and economic development will be proposed.

STATIC ANALYSIS OF CREATIVITY AND THE TECHNOLOGY OF THE FIRM

At the level of the firm, creativity has two effects

1. it influences the combination of inputs to generate an output, and
2. it influences the combination of inputs so as to change the characteristics and thus the unit value of the output.

The first type is often called process R & D, while the second is called product R & D.

Both types of creative activity are normally integrated. Decisions on the two types of creative efforts can be illustrated by model (5):

$$\text{maximize} \quad \pi = p(K_1)\ Q(C,K_2) - pC - \psi_1 K_1 - \psi_2 K_2; \tag{5}$$

where $\pi \doteq$ profit; $p \doteq$ product price; $Q \doteq$ flow of product; $C \doteq$ material capital; $K_i \doteq$ knowledge of type i; $p, \psi \doteq$ factor prices.

Necessary conditions of a maximum are:

$$\begin{aligned} (\partial\pi/\partial K_1) &= (\partial p/\partial K_1)Q - \psi_1 = 0 \quad ; \\ (\partial\pi/\partial K_2) &= p(\partial Q/\partial K_2) - \psi_2 = 0 \quad ; \\ (\partial\pi/\partial C) &= p(\partial Q/\partial C) - p = 0 \quad ; \end{aligned} \tag{6}$$

implying that there are economies of scale in product development, because the cost of improving the product is spread over all units produced.

In the sequel, we disregard the distinction between process- and product R & D by formulating a value function F.

ENDOGENOUS TECHNOLOGICAL DEVELOPMENT IN ECONOMIC MODELS

A two sector model

Little has been written on endogenous technological development since the days of Schumpeter.

In an article in 1978 I proposed a neo-classical growth model, in which technological development was determined by R & D-investments taken out of value added {Andersson, 1978]. The simplest version of the model has only one sector of production and one R & D-sector:

$$\dot{C} = [1 - \mu]\, \sigma\, F(C,K) \quad ; \underline{\text{Investment}}$$
$$\dot{K} = \gamma\, \mu\, \sigma\, F(C,K) \quad ; \underline{\text{R \& D}} \tag{7}$$

where $K \doteq$ stock of knowledge (proportional to service flow)

$\mu \doteq$ investment allocation ratio for non-material (or R & D) purposes

$\gamma \doteq$ productivity of the R & D-sector

$C \doteq$ stock of material capital.

Labor is here assumed to be pre-allocated between the activities of the economy. Endogenous determination of labour use does not significantly alter the characteristics of the model [Andersson, 1978] .

If γ, μ and σ are parametrically given and F is well-behaved in the neo-classical sense, then a general equilibrium growth trajectory can be proved to exist. Furthermore, it will be relatively stable, in the sense of Nikaido [1970].

The knowledge-material-capital-ratio will, in the long run, converge towards a constant, determined by the investment allocation ratio, μ, and the R & D productivity, γ.

Optimal control of μ of the system can also be deduced. The optimal level of μ depends on the exact form of the production function and the productivity of the R & D-sector (and the goal function).

For goal functions, independent of μ, one invariant result emerges: the shadow-prices of material capital and knowledge should be kept at a constant ratio, equal to the productivity of the R & D-sector. With a homothetic production function, this implies that the system expands along the turnpike with constant proportions of knowledge and material capital.

Multisectoral Economy-Technology Models

The neo-classical two-sector model of endogenous growth of knowledge is a convenient starting point (but a starting point, only) for a more general study of creativity and economies.

One extension is into a multisectoral framework, with one sector producing commodities and many sectors producing knowledge. In this way, the complication of trade between sectors is avoided. A possible extension of the former neo-classical model is the following system of differential equations:

$$\begin{aligned}
\dot{C} &= \sigma(1-\mu_1-\mu_2-\mu_3)\, F(C,K_1,K_2,K_3) \\
\dot{K}_1 &= H_1(\sigma\ \mu_1\, F(C,K_1,K_2,K_3);K_1,K_2,K_3) \\
\dot{K}_2 &= H_2(\sigma\ \mu_2\, F(C,K_1,K_2,K_3);K_1,K_2,K_3) \\
\dot{K}_3 &= H_3(\sigma\ \mu_3\, F(C,K_1,K_2,K_3);K_1,K_2,K_3)
\end{aligned} \tag{8}$$

with $\sum_{j=1}^{3} \mu_j = 1$ and $\mu_j \geqq 0$

$H_j \geqq 0$, continuous, and monotonously increasing with σ, μ_j and $\{K_i\}$,

where $K_j \doteq$ stock of knowledge of type j, assumed to be real valued and augmented by R & D of type j($\doteq \dot{K}_j$).

Every production system is located on some <u>organizational network</u> with some frictions of communication between nodes, separated by links.

Accessibility measures have been proposed to represent the effect of such frictions on the value of non-tradeable resources. Fujita and Smith (1985) have shown that a discounting formula representation of accessibility is equivalent to a contact maximizing search procedure:

$$A = \sum_{j=1}^{3} K_j \exp\{-\beta d_j\}$$

where $d_j \doteq$ organizational "distance" to knowledge of type j.

A spatial version of system (8), employing the given accessibility representation, has been studied in Andersson-Mantsinen (1981):

$$\dot{C}_i = \sigma_i(1-\mu_1-\mu_2-\mu_3)\ C^{\alpha}{}_i(\Sigma\ K_j\ \exp\{-\beta d_{ij}\})^{\kappa} \qquad (9)$$
$$\dot{K}_i = \gamma\mu_i\ C^{\alpha}{}_i(\Sigma\ K_j\ \exp\{-\beta d_{ij}\})^{\kappa}$$

Assumption $\mu_i > 0$, $\Sigma\ \mu_i = 1$ $(i = 1,2,3)$

$\alpha > 0$; $\beta > 0$; $\kappa > 0$; $C_i(0) > 0$; $K_i(0) > 0$; $\sigma_i < 1$; $\gamma > 0$.

Extensive simulations with model (9) under different parameter assumptions, within the bounds given, have shown that:

* the system converges to some positive common growth rate in the long run, as expected by the theorem due to Nikaido (1968),
* increasing σ_i increases the common long term equilibrium growth rate,
* decreasing any distance, d_{ij}, increases the common long term equilibrium growth rate,
* changing the parameters μ_i from a slow growth into a fast growth pattern leads to steady convergence of the growth rates of different sectors towards the new, higher common long term equilibrium rate of growth if $\alpha + \kappa < c^*$. If $\alpha + \kappa > c^*$, growth rates increase much faster in the central nodes than in the periphery of the network, leading to increased inequalities between center and periphery.

SYNERGY AND CREATIVITY

It has been stressed by numerous analysts, e.g. Weidlich, Mensch, Haag [unpublished manuscripts] , building on Haken(1980), that the development of knowledge should be seen as a synergistic process, i.e. a process in which interactions between variables of a system are essentially non-linear. Model (9) can be rephrased so as to capture the argument.

$$\dot{C} = \sigma(1-\mu_1-\mu_2-\mu_3)\, C^{\alpha}(\sum_j K_j \exp\{-\beta d_j\})^{\kappa} \tag{10}$$

$$\dot{K}_1 = \sigma\mu_1 C^{\alpha}(\sum K_j \exp\{-\beta d_j\})^{\kappa}[a_1+a^1{}_{12}K_1K_2+a^1{}_{13}K_1K_3+a^1{}_{23}K_2K_3+a^1{}_{123}K_1K_2K_3]$$

$$\dot{K}_2 = \sigma\mu_2 C^{\alpha}(\sum_j K_j \exp\{-\beta d_j\})^{\kappa}[a_2+a^2{}_{12}K_1K_2+a^2{}_{13}K_1K_3+a^2{}_{23}K_2K_3+a^2{}_{123}K_1K_2K_3]$$

$$\dot{K}_3 = \sigma\mu_3 C^{\alpha}(\sum_j K_j \exp\{-\beta d_j\})^{\kappa}[a_3+a^3{}_{12}K_1K_2+a^3{}_{13}K_1K_3+a^3{}_{23}K_2K_3+a^3{}_{123}K_1K_2K_3]$$

A second order interaction is given by $a^h{}_{ij}K_iK_j$, a third order interaction is given by $a^h{}_{ijl}K_iK_jK_l$, etc.

System (10) cannot be expected to have any simple analytic properties. It is, however, well known that systems of this kind can behave in rather unexpected ways at some patterns of parameters. It should thus be expected that, if the system is stable in some development phase, it can easily "flatten out" and be unstable and even unpredictable in some other phase, when the slowly changing parameters $(a^h{}_{ijm})$ have gone through some critical threshold.

It is in these phases that new knowledge development plans should be formulated. Qualitative jumps of the system (10) can be expected to occur in these stages, even if the shifts of μ_j-controls are small.

The problem of inefficiency of decentralization of R & D decisions occurs even without the introduction of synergistic interaction. In this case we have $F_i(C_i,K_1,K_2,K_3,\dots,K_i,\dots) = C^{\alpha}{}_i K_1\exp\{-\beta d_{i1}\} + C^{\alpha}{}_i K_2 \exp\{-\beta d_{i2}\} + \dots C^{\alpha}{}_i K_i \exp\{-\beta d_{ii}\} + \dots$ if $\kappa = 1$. Assuming output price equal to unity, profit maximization of sector i requires $\partial F_i/\partial C_i - p = 0$ and $\partial F_i/\partial C_i - \psi = 0$, where p and ψ are the prices of material capital and knowledge, respectively. This is not an optimal position for the system as a whole, because K_i enters all the F_j-functions where $(j = 1,\dots,i,\dots,N)$. The larger the N, the larger the degree of inoptimality, cet.par. Synergetic effects <u>reinforce</u> this conflict between decentralized optimality and system optimality.

THE ROLE OF SYNERGIES

The consequences of synergies are currently under numerical analysis based on system (10).

Some empirical observations of the role of synergetics have been made in association with historical studies of creative regions. Examples of such regions are:

1. Athens 500-300 B.C.	
2. Florence 1400-1500	Andersson [1985]
3. Vienna 1880-1930	Janick & Toulmin [1978]
4. New York 1950-1980	Hoover and Vernon [1961]
5. San Francisco 1960-	Andersson [1985]

In all of these regions of creative explosions, the following characteristics have been recorded:

1. A starting point of deep knowledge in a number of scientific and artistic fields.
2. A sponsoring institution with a tolerant attitude to different scientific and artistic activities.
3. A perceived social disequilibrium.
4. Possibilities of planned and spontaneous intensive local interaction and extensive non-local interaction.
5. Institutional and other structural instability or genuine uncertainty.

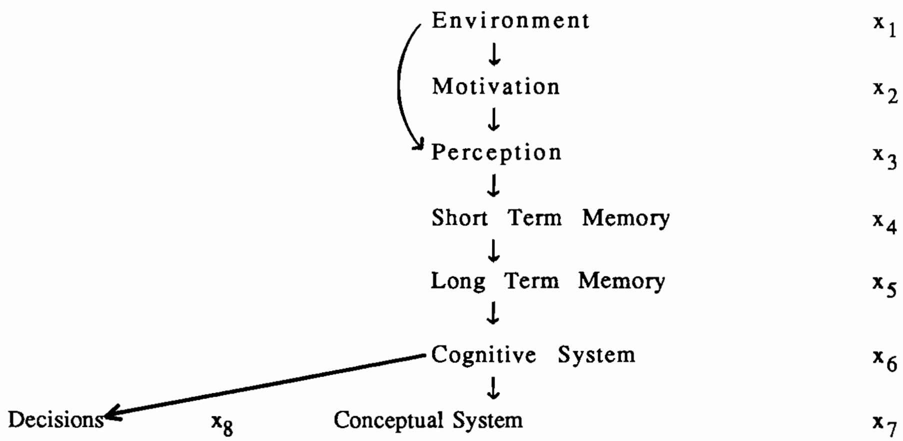
These observations lend some credibility to the modelling of creativity, technological and economic development as a synergistic process.

A fruitful theory of creativity can not, however, be based on a macro foundation only. A micro theory is needed. As shown above, not much has to be added to the neo-classical theory of production to generate a theory of decisions on employment of creative persons, provided the firm is big enough to handle the choices deterministically. When we require a micro theory of creativity, it is not a quest for a new theory of the firm.

MICRO ANALYSIS OF CREATIVITY

Creativity is not only a social phenomenon, occurring in organizations, firms and regions. In many cases, creativity must be seen as a dynamic psychological phenomenon, associated with individuals. To understand creativity at this level, we need a fundamentally dynamic theory of the cognitive process, i.e. to paraphrase Polya, not only a theory of "problem solving" but also a theory of "problem finding".

Psychologists normally model the brain as a basically linear recursive system with the following components.



x_i are vectors of variables often assumed to be real-valued.

Figure 2 Classical Recursive Structure of the Psychological System

It is usual among economists to accept this model in a condensed version:

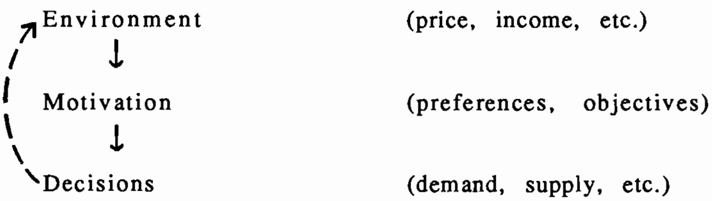


Figure 3 Condensed Recursive Psychological System

Two major advantages of the classical paradigm are obvious:
With the recursive formulation, the standard experimental strategy of psychology is facilitated. It is quite easy to control the experiment, if all relations are recursive.

Statistical estimation is also facilitated by a recursive model formulation. It has been shown (Wold, 1964) that ordinary–least–squares methods can be used in recursive systems without identification problems and other losses of statistical qualities.

Unfortunately, recent studies of human and artificial intelligence have undermined the basis of this handy formulation of the classical, recursive and linear psychological paradigm.

The Perception System

A recent study by Stewart and Peregoy (1983) shows that the perception system must be modeled as a non-linear dynamic subsystem. With series of experiments, employing a standard perception stimulus, they have shown that the stimulus response relations are best represented by a catastrophe model. The model of perception can be illustrated by two diagrams.

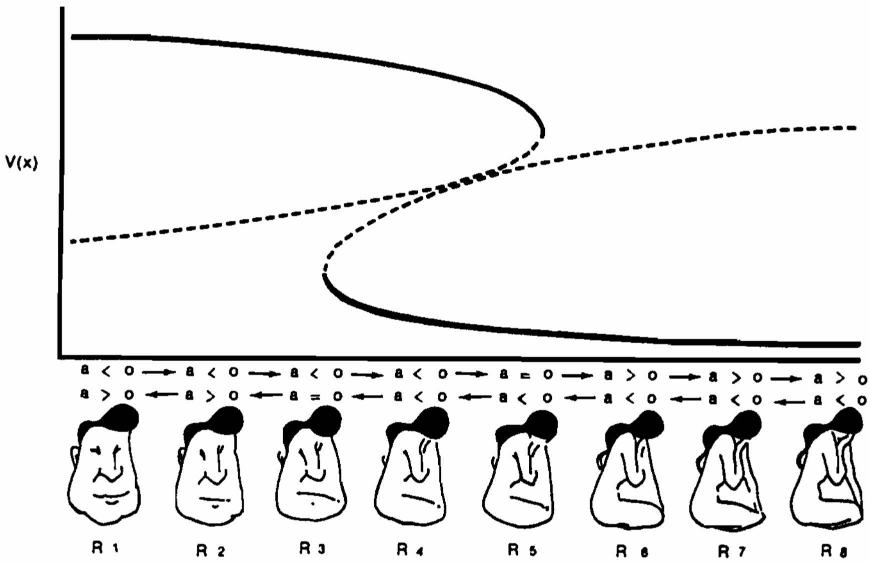


Figure 4 Two Fold Catastrophes Developed by Beginning the Experiment at two Different Ends of Stimulus Sequence with Hysteresis

Despite some differences in patterns of response among different groups of subjects, the result is clear: for the vast majority of subjects it is not possible to represent their perception structure by a linear, dynamic model.

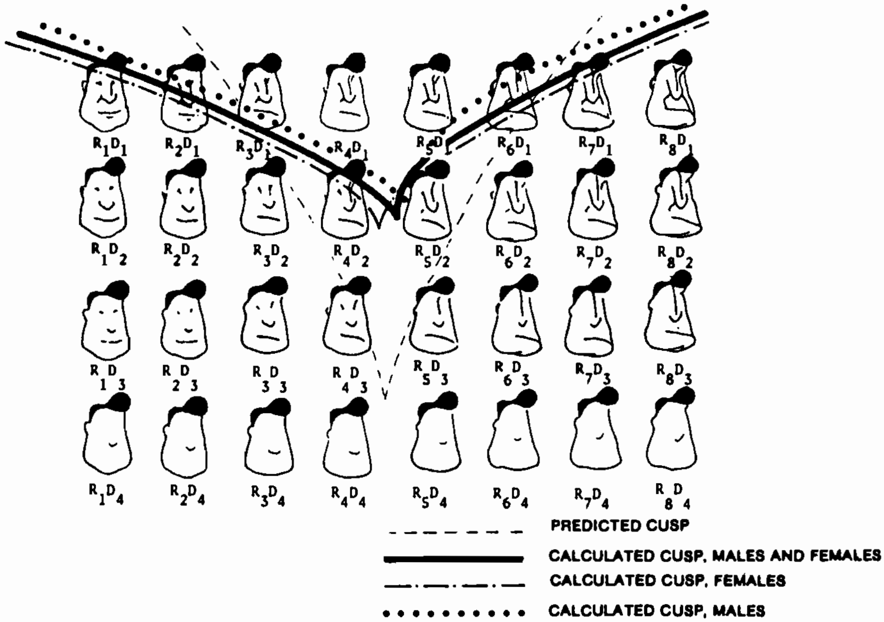


Figure 5 Estimated (calculated) Cusps for Different Groups of Subjects in Stewart-Peregoy Study

The Cognition-Recognition System

Pattern cognition and recognition can mostly be regarded as problem solving. Any child permanently recognizes his or her mother and other relatives from the age of a year or two. Most computers are quite inefficient, however, so there seems to be some basic logical difference between the artificial and human cognitive systems. In a few – not well controlled – experiments, I have studied the response to a meaningful figure from Hofstadter (1980), shown in Figure 6.

Figure 6

The respondents were asked to look at the figure and respond directly to:

1. what they recognized, and report
2. what they could see, when some other interpretation eventually became possible.

The results were the following.

The majority of respondents answered that the immediate impression is a piece of abstract art. After a time delay of 30 to 180 seconds the response changed and an intellectually meaningful pattern was reported. There were, however, great differences in delay time between people with different conceptual systems.

The experimental group consisted of 30 musicians and 15 artists. The artists were generally faster in jumping from the first, artistic, to the second – intellectual – interpretation. Training, i.e. conceptual development, obviously enhances the cognition-recognition pattern.

The experiment shows that at least a fold catastrophe model is needed to represent the non-linear nature of the cognitive system.

The Memory System

Traditionally, the brain has been assumed to be an inventory with located storages. This inventory is then often assumed to have two sequential compartments. The first compartment – "the short term memory" – is subjected to rapid depreciation and only sends on a limited amount to the second compartment – the "long term memory", where items to be remembered are assumed to be stored in well located sites for later retrieval. This static and spatial theory of the memory is now being substituted for a dynamic, non-spatial theory [Nilsson (1984), (1985)].

In these new models, the stress is on memorizing rather than memory. Remembering is seen as an interaction between a set of available cognitive capabilities in situations when an individual is motivated to remember something. Information is assumed to be distributed over the whole system without retrieval addresses. To quote Nilsson (1985):

> "The basic argument to be made with respect to storage of information, is that the whole storage system undergoes a subtle change whenever an encoding takes place. Thus, it is proposed that no single memory trace is formed at the time of encoding. Instead it is argued in line with Craik [4] "that the system has an increased likelihood of recreating the same pattern of activity on a subsequent occasion, especially if many aspects of the original event are re-presented to help drive the system into the same general configuration" (p. 345).
>
> The reasoning behind this is as follows. It is assumed that situations, objects, pictures, words etc., constituting the to-be-remembered (TBR) information, and the context in which these items occur can be decomposed into a large number of physical features. These features are assumed to furnish the individual with the basis for various perspectives or "affordances" [6], for each particular object, word etc. The information about a given situation is there to be picked up by the individual and different combinations of features are assumed to set the stage for potential affordances.
>
> The current cognitive environment [17] of the individual as invoked by instructions, current mood states, and motivation is assumed to determine which points will be activated by means of this spreading activation of interrelated points. Aspects combine to form what we refer to as functional dispositions [18] .
>
> A functional disposition serves as a means for acquiring the information conveyed by the current situation, that is for integration of this new information with the overall, already existing experience and knowledge of the individual. This disposition is conceived of as a psychophysiological process, which increases the neural excitability of those points activated. This increase in neural excitability serves the purpose of facilitating the formation of new neural connections. Expressed differently a function disposition is a wider and more complex activity pattern than an aspect, and the functional disposition means a readiness in its most global sense for the individual to act in accordance with

what the current object or event has to offer. At any given moment the rememberer is prepared to act, or in other ways deal with the situation, on the basis of the affordances conveyed by the object and the functional dispositions activated."

Source Nilsson, (1985).

The conclusion of the analysis is that remembering is not separable from cognition and conceptualization. Remembering becomes an integral part of the cognitive process, earlier shown to be a non-linear dynamic process. The essential factor in remembering is the activation of conceptual models, organizing the perceptions of the environment.

A SYNTHESIS

A micro theory of creativity can be tentatively formulated as a reflection of our discussion of the perception, cognition and memorizing systems. This system can be seen in the following abstract way:

$\dot{x}_1 = f_1(x_1, x_2, x_8, \epsilon)$	<u>Environment</u>
$\dot{x}_2 = f_2(x_1, x_2, x_6)$	<u>Motivation</u>
$\dot{x}_3 = f_3(x_1, x_2, x_3, x_6, x_7, x_8)$	<u>Perception</u>
$\dot{x}_6 = f_6(x_1, x_2, x_3, x_6, x_7, x_8)$	<u>Cognition and Memorizing</u>
$\dot{x}_7 = f_7(x_1, x_2, x_3, x_6, x_7, x_8)$	<u>Conceptualization</u>
$\dot{x}_8 = f_8(x_1, x_2, x_3, x_6, x7, x_8)$	<u>Decisions</u>

Memorizing is part of the solution of this non-linear dynamic system. We can now define a creative micro-process to be one in which this indecomposable dynamic system has an expanding conceptual system i.e. with $\dot{x}_7 > 0$. This does not preclude an equilibrium. It is obviously possible to have a situation where $\lambda x_i = \dot{x}_i$, all i, with $\lambda > 0$, a situation of general learning.

Because of the essential non-linearity of the perception and cognition system there can be no uniqueness of solutions of the total system.

As an alternative to this equilibrium approach to the analysis of cognition, G. Hinton has designed a model based on repeatedly minimizing energy or neg-entropy, until an optimum which is lower than other reachable minima is found. Thus, at least heuristically, some of the problems of non-uniqueness of solutions are avoided. However, the Hinton approach makes the implicit assumption that decisions of the creative process are triggered by a wish to avoid use of energy, a disputable motivational assumption.

Most of the evidence indicates that external motivations (economic rewards and other inducements) have little effect on the general creativity of individuals, given the choice of education, region, and occupation (Amabile, 1983, Simonton, 1975, Smith, Carlsson and Danielsson, 1984). Environmental stimuli primarily influence the choice of education, occupation and region, which indirectly influences the intensity use of an individual creative potential.

INTEGRATING MICRO AND MACRO ANALYSIS OF CREATIVITY

In the same way as psychological decision analysis and engineering production theory have been built into static economic market models, the micro analysis of creativity must be integrated into the macro analysis.

It is clear from our discussion above, that the brain operates in a non-linear, often structurally unstable way in perceiving, solving or finding problems as well as remembering. Slow and steady changes of the environmental parameters can trigger off completely new steady state solutions in terms of perceptions and conceptual elements like theories, models and methods.

Empirical evidence suggests that creative societies are normally confined to small, integrated environments, well connected to other similarly small integrated environments. Macro analysis should be based on this assumption. It means that small dense networks of non-linear micro systems are connected into a large network of communicating creative organizations or regions. The implication of this is that the synergisms of type (10) models could be deduced from analysis of small integrated creative organization models, in which smallness makes the micro units distinguishable.

CONCLUSIONS

After a long period of analysis of economic adaptations to parametric changes of technology, economics could move to an endogenous treatment of economic and technological interactions. Some macro models based on the Schumpeter (1934) paradigm have been advanced. However, these models suffer from a lack of creative activity analysis at the micro level.

A micro model of creativity is advanced. It is argued that such a micro model must be non-linear and dynamic 1) . Furthermore, such a model should not be expected to have the property of uniqueness of solutions. Rather, non-uniqueness or structural instability is the key to an understanding of creativity at the micro level (as well as – for other reasons – at the macro level).

Macro analysis of creativity must take into account the empirical evidence suggesting that creative environments are composed of synergistic groups of creative individuals on dense parts of communication networks.

NOTES

1) It is quite possible that the differential equation paradigm is too simple for the purpose. See Rosen (1985) for an alternative.

REFERENCES

Amabile, T.M., The Social Psychology of Creativity, Springer, Heidelberg, New York, 1983.

Andersson, Å.E., 'Growth and stagnation of economics with public goods: a neo-classical analysis', Lectures on Regional Stagnation, Buhr, W., Friedrich, P. (eds.), Nomos, Baden-Baden, 1978.

Andersson, Å.E., Mantsinen, J., 'Mobility of resources, accessibility of knowledge, and economic growth', Behavioral Science, 25:5, 1980, pp. 353-366.

Andersson, Å.E., Kreativitet - Storstadens framtid, Prisma, Stockholm, 1985.

Fujita, M., Smith, T.E., Unpublished manuscript, University of Pennsylvania, 1985.

Haken, H., Dynamics of Synergetic Systems, Springer, Heidelberg, New York, 1980.

Hofstadter, D., Gödel, Escher, Bach, An Eternal Golden Braid, Vintage, New York, 1980.

Hoover, E.M., Vernon, R., Anatomy of a Metropolis, Harvard University Press, Cambridge, Mass., 1959.

Janik, A., Toulmin, S., 'Wittgensteins Vienna', Cambridge University Press, New York, 1978.

Nikaido, H., Convex Structures and Economic Theory, Academic Press, New York, 1968.

Nikaido, H., Introduction to Sets and Mappings in Modern Economics, North Holland, Amsterdam, 1970.

Nilsson, L.G., 'New functionalism in memory research', in Lagerspetz, K.M.J., Niemi, P. (eds.), Psychology in the 1990s, North Holland, Amsterdam, 1984.

Nilsson, L.G., 'An interactionistic approach to models of distributed memory', Umeå, forthcoming

Rosen, R., Anticipatory systems, Pergamon Press, London, 1985.

Schumpeter, J.A., The Theory of Economic Development, Oxford University Press, New York, 1934.

Simonton, D.K., 'Sociocultural context of individual creativity', Journal of Personality and Social Psychology, 32, 1975, pp. 1119-1133.

Smith, G., Carlsson, I., Danielsson, A., 'Experimentella undersökningar av kreativitet', Psykologiska institutionen, Lund, 1984.

Solow, R., 'A contribution to the theory of economic growth', Quarterly Journal of Economics, 70, 1956, pp. 65-94..

Stewart, I.N., Peregoy, P.L., 'Catastrophe theory modeling in psychology', Psychological Bulletin, 94:2, 1983, pp. 336-362.

Swan, T.W., 'Economic growth and capital accumulation, Economic Record, 32:63, 1956, pp. 334-361.

Wold, H., (ed.), Econometric Model Building, North Holland, Amsterdam, 1964.

Chapter 3

The General Theory of Disequilibrium Economics and of Economic Evolution

R.H. DAY

> Existence is either ordered in a certain way, or it is not so ordered, and conjectures which harmonize best with experience are removed above all comparison with conjectures which do not so harmonize.
>
> THOMAS HARDY

Various kinds of simple dynamic economic behavior are well understood: the existence and character of stationary states, steady or balanced economic growth, and periodic business cycles. Each of these types of behavior has its corresponding explanation or set of alternative explanations. Theories of general equilibrium explain stationary states or steady, balanced growth. Theories of business cycles explain periodic oscillations in the economy. Unfortunately, simple dynamic behavior is not exhibited by typical economies of record. Instead they exhibit complex dynamics: irregular fluctuations; overlapping waves of development and structural change; and institutional change and evolution.

If there were a tendency for economies to converge to simple dynamic paths within a fixed institutional framework, none of this would be important, because the departure from balanced growth or cycles would eventually abate. Theories of the steady state and of cycles would approximate with ever greater accuracy the path of actual events, and society would settle down once and for all to a fixed organizational structure. But this too is not the case. If anything, the pace of change has accelerated with the advance of human progress; the duration of growth and decay periods have shortened correspondingly. Fluctuations have dampened for a time, only to erupt again in even wider swings; in spite of the remarkable growth in statistical methods of estimation, progress in forecasting is negligible at best. Economic change is as erratic, or even more so, than ever.

The ubiquity of complex behavior has not dissuaded theorists from extending the theory of equilibrium from the mere explanation of stationary states and balanced growth to a rationalization of the business cycle. This is not because they eschew an interest in the empirical facts of change. Indeed, some of the most beautifully motivated and influential work in this direction has been aimed precisely at the explanation of the stylized facts of the business cycle. But to square with reality, it has been necessary to augment the equilibrium theory and its underlying convergence postulate with an assumption of perturbing exogenous shocks. These presumed shocks will be seen to fall in two classes. In one class are structural changes (foreign aid, migration, technological discovery) sufficient to induce the observed waves of development. In the second are more or less random shocks - such as weather, political tampering with policy variables, earthquakes, etc. - which continually perturb economic motion and give it the irregular character universally observed in the data. Given their presence, the facts of complex change are quite in accord with equilibrium and the convergence postulate.

From the point of view of pure theory, this is an unsatisfactory state of affairs, however, because it rests on two ad hoc assumptions: that of convergence and that of exogenous shocks. Without denying the practical necessity in empirical work of incorporating exogenous variables, and even of assuming the perturbing influence of random shocks, there is still the open question as to how an economy would behave when the former are constant (or some other simple function of time) and when the latter are absent altogether. If the convergence postulate does not hold, then complex dynamics may very well persist; and whether or not it holds and under what conditions is an open question.

Somewhere Paul Samuelson observed that economic equilibrium is a state which, if brought about, would have certain properties. His colleague Frank Fisher, in a cogent review of the literature, pointed out the deficiencies in the theory of how such equilibria might be reached. He went on to consider a disequilibrium foundation for equilibrium economics in an attempt to alleviate the theoretical lacuna. But I think it would be more instructive - as it often seems to be the case - to turn his sequence around, and to recognize in equilibrium a foundation for a disequilibrium theory of economic change. This is because it, in effect, asserts that if an economy is out of equilibrium it must change, and of course, by hypothesis, it must change out of equilibrium. This leads to a modelling problem: how do economies work in disequilibrium? And this leads to several analytical questions: Under what conditions will an economy in disequilibrium converge? When will it perpetuate change in disequilibrium? To this we may add, when will that change be complex? In the following pages, the beginnings of answers to these questions are outlined.

CHANGING ECONOMIES

Let us begin by identifying salient features of actual economies that must be the basis for any theory of change. Consider that time is decomposed into elemental periods (days, weeks, months, seasons, years). At the beginning of each such period, a state of technology, resource availability, social organization, and individual preferences prevail, and, of course, a history of past consumption, production and technological practice has occurred. On the basis of this, individuals and organizations in the economy make their plans, modifying or retaining old plans or drawing up altogether new ones, and carrying out various other actions. In the next period the situation has changed. Resources have been depleted, capital may have been augmented, prices and other indexes of value and wealth will be modified and so on. The system is poised for a new round of planning and action.

Observed over a sequence of periods, the economy will exhibit a history of specific activities that were and were not pursued, of specific technologies and resources that were and were not utilized, of specific constraints that were or were not binding. In the course of this process the consumption and/or production activities actually utilized change; i.e. the constraints actually impinging on choice and actions switch. Some variables that appeared relevant will no longer appear so; other variables that once seemed of no importance at all will now appear to play an active role in development; some technologies may be abandoned, different ones taking their place; some resources, previously plentiful and perhaps thought of as free goods, now become scarce and attain great value in exchange; still other resources, once crucial in the production transformation, are abandoned, perhaps even before they are exhausted, again becoming valueless.

Viewed in the aggregate, waves of growth or decline in productivity and fluctuation in output and value will occur, and in the long run, various "epochs" or "ages" will appear, dominated by characteristic activities and resources. In the short run, we shall see individuals and organizations occasionally change what they do and how they do it. In general the economy's technological regimes will switch, its consumption and production patterns will change: its structure and behavioral patterns will evolve.

It will not appear to converge to states that have the earmarks of equilibrium. Individuals will rarely be seen to do their best; they often experience regret, and from time to time they are forced to change their plans or even to act in a manner contrary to plan. Markets rarely balance, and individual plans are often incompatible. Normally, some people are becoming better off, while others are becoming worse off, frequently the former at the direct expense of the latter. Neither do aggregate indexes of activity converge to steady states of balanced growth or circular flow.

To summarize: economic agents rarely achieve optimal plans; their plans are often inconsistent; the flows their actions generate are out of balance; their fortunes fluctuate in divergent paths; the economy is a disequilibrium process.

THE ECONOMY AS AN ADAPTING PROCESS

A fruitful starting point for a theory of the evolving, disequilibrium economy is the adaptive process, a dynamic system in which a behavioral unit of interest, the agent, (firms, households, government bureaus, and individuals in them) responds to its own internal conditions and to prevailing circumstances in its environment. Since agents and environment influence each other, interactive feedback is involved. From each agent's point of view, other agents are part of the environment. The economy may then be thought of as being made up of a set of interacting adaptive processes, i.e., as a complex, adapting system.

To be feasible, action must be consistent with internal states, but since the intended actions of a given agent may be inconsistent with those of other agents in the environment, it may be impossible to execute them. Therefore, actions must generally be based, not just on intentions, but on internal states and/or external states. These have the effect of insulating the agent from inconsistencies with the outside, or make it possible for the agents to generate unbalanced flows (consumption exceeds production or vice versa, etc.). This must happen in either of two alternative ways. One is the stock-flow mechanism. The other is the contingent tactic.

STOCK-FLOW MECHANISMS

Actions bring about material, financial, informational, and energy flows which modify internal and external stocks. The flows among various agents, based on individual adaptive processes, are in general imperfectly coordinated. The resulting imbalances in flows are mediated by stocks, which make it possible for flows into a given agent to be unequal to flows out from the given agent. Internally, the agent can maintain strategic reserves of materials and energy potentials to enable production and consumption to take place if shortages or delays in supply occur. And they can maintain financial reserves (cash balances, liquid portfolios) to make possible a flow of expenditures that might otherwise be curtailed when sales fall.

Externally, special agents or institutions whose primary function is to mediate unbalanced flows by regulating stocks have come into being. For example, stores are inventories on display, which make it possible for consumers to purchase goods, and producers to supply them, without either knowing the plans or actions of the other. Banks and other financial intermediaries regulate the flow of purchasing power from uncoordinated savers to investors. Their ability to create credit provides a means of coordinating activities at different points in time, and of facilitating exchange when current monetary stocks are inconsistent with intended investment or consumption expenditures. Of course, there is nothing in what we have said so far which guarantees that existing internal or external stocks will succeed in providing the buffers required, or, if they do for a time, that they will continue to do so. That is why another possibility is pursued when volition can be exercised.

INTENDED ACTION AND CONTINGENT TACTICS

In addition to stock/flow adjustments, which have the effect of rendering feasible the unbalanced flows induced by disequilibrium actions, there is the contingent tactic, which facilitates a timely remedy for intended, but currently infeasible actions. Generally speaking, the process of generating intended actions is more or less elaborate and deliberate. It takes time, uses resources, and is therefore costly. But actions must take place continuously, and in order to avoid a catastrophic crisis of inaction, contingent tactics must be timely: they must not take too much time to formulate and they must not use too many resources to execute. And of course they must work. But by the very nature of disequilibrium and its inherent uncertainty, viability is not guaranteed, and in "the real world" there is evidence on every hand that it is not always achieved.

MODES OF ECONOMIZING

Our concern is with economic activity including resource allocation, production distribution (including exchange) and consumption. The process that generates behavior of this kind is economizing behavior because in general it requires the use of scarce resources and involves a trade-off of alternative ways of doing things. Economizing, in this sense, is a mental activity leading to economic action. To accord with the facts of disequilibrium we have just outlined, it must incorporate a crucial distinction between intended actions and contingent tactics. Consider now how intentions and contingent tactics of economizing behavior are formed.

As we have seen, economizing takes place within a complex, adapting system based on stock-flow adjustment mechanisms. It is constrained by the inherent limitations of the mind. There is the imperfect perception and knowledge of the environment. Existing states are perceived imperfectly and the feedback structure that determines how the environment of a given agent works is only partly understood. The exercise of conscious thought involves limited memory, limited recall and limited powers of ratiocination. These imply limits on the ability to solve complex decision problems. Moreover, individuals do not always know what they want; their preferences are incomplete or undefined. Still further, individual people possess limited capacities for interpersonal communication and cooperation, frailties which are amplified in organizations, a fact illustrated in the childhood game of "telephone".

In short, both thought and communication require effort, take time and are imperfect.

BOUNDEDLY RATIONAL CHOICE

Rational choice is the conscious, deliberate process of selecting the most preferred among perceived alternatives. As a means of arriving at intended actions and contingent tactics, it is bounded by cognitive limits and is imperfect. Faced with cognitive limits, individuals form simplified representations of alternative activities and constraints. Their optimization is therefore proximate. In routine situations they conceive of choice as a departure from previous activity, and explicitly consider a small number of alternatives in the neighborhood of what is familiar. The willingness to depart from current practice - that is, the extent of the region searched - may depend on experience and on the behavior of other agents. Thus, adaption to current economic opportunity may be more or less flexible. The set of alternatives that may be considered at a given time, called the zone of flexible response, depends on experience and imitation. This dependence means that economizing is both adaptive and more or less cautious.

The choice within the constraints determined by technology, resource availability, and by the willingness to be flexible in responding to opportunity, is directed by preferences which are represented by various goals which may be arranged according to some (perhaps temporary) hierarchy or priority order. A first goal dominates comparison of alternatives until a satisfactory solution is obtained according to this goal; then a less important goal is used to choose among the alternatives satisfying the higher order goal, and so on, until a single choice is reached.

OBEDIENCE, IMITATION, HABERATION, EXPERIMENTATION

Rational thought requires effort and takes time and resources itself, and it can only be effected when well-defined preferences exist. But there are other options. These include obedience to an authority, that is, doing what you are told; and imitation, that is, doing what someone else is doing or has already elucidated. Both of these may be attractive models of behavior compared to thinking for one's self, because they save intellectual effort for other mental tasks and because they make accessible possibly superior forms of behavior that could not have been created through one's own exercise of imagination and rationality.

An additional mode is universally involved in human economizing activity. It is to do what one has been doing. This allows for a kind of mechanical, unconscious mode of behavior that requires neither imagination or rationality, and is thus even more stingy with mental energy than imitation. It enables one to behave according to a habitual pattern. Because the English language does not contain a verb meaning to act according to habit, I coined the term habere (habear). If I habere, I execute a frequently repeated sequence of actions that requires little if any conscious thought. "Habering" or "haberation" is certainly an extremely important mode of behavior, and in a mind of bounded rationality, an indispensible faculty for economizing the mental energy that drives conscious thought.

Still another mode of behavior must be distinguished. This allows for purposeful activity when the conditions of rational thought do not exist, when a habit appropriate to the purpose is yet unformed, or the motivation for imitation is lacking. As a general mode it may be called experimentation. It may involve a systematic exploration of a controlled environment or model as a way of arriving at a decision. It might involve trial and error search in a sequence of local experiments, in which the direction of search is modified in response to a measure of success or failure. Or, it may involve vaguely purposeful exploration or even play. It can be directed at solving all sorts of mental and physical problems or it can be essentially unmotivated. In either case, it is a freewheeling, sometimes more or less random process that involves trial of alternative thought or action patterns when careful methods of ratiocination cannot be exercised, or when the requisite skills have not been acquired.

THE ECONOMY OF MIND

All of these modes play a useful role in the allocation of scarce intellectual capacity to alternative purposive tasks. They imply the existence of a higher level cognitive faculty that must direct the mode of mental activity to that governing behavior at any given time. What is implied in this description of behavior is an economy of mind: a system of mental resource allocation and of choice among alternative modes of transforming internal or mental states and information about the external world into economic choice and effective behavior. Such a higher order faculty cannot operate according to the usual laws of pure economic rationality, however, because the consequences of choosing one over the other mode of behavior are rarely known. At the risk of introducing a confusion with other uses of the term, I shall call this faculty judgement. How it works is a matter which should be of concern to economists, for its exercise must be a routine aspect of economic behavior. It is responsible for orchestrating a system of information processing, planning and control that will lead to intended action which is practical, i.e., which can be realized as often as possible. In addition, it leads to a contingent tactic, or hierarchy of contingent tactics that can take over the governance of economizing behavior when intended actions are infeasible. Such algorithms will involve one, or more, or perhaps even all of the economizing modes we have mentioned. Since all of them involve internal and environmental feedback to the agent, we may refer to such a system as adaptive economizing with feedback.

RECURSIVE PROGRAMMING AND MULTIPLE-PHASE DYNAMICS

The mathematical analysis of such a dynamic, multi-mode microeconomic theory has scarcely begun, but one example emphasizing boundedly rational economizing with feedback has been extensively applied. This is the recursive programming model, a dynamical system in which behavior is represented by cautious, local optimizing subject to stock and flow constraints, to constraints that define the local region of flexible response. Here the constraints depend recursively on past behavior of the agent and other agents in the environment in a way that represents the accumulation of stocks, and the effects of imitation and haberation. The solution of such a model typically exhibits changing modes of behavior, nonperiodic fluctuations and sensitivity to perturbations in initial conditions and parameter values. In addition, they exhibit changing sets of utilized activities and tight constraints. When these sets switch, the variables and equations governing the evolution of the system switch, in effect bringing in a different set of causal structures and feedback loops. These structures are called phase structures of regimes. A given model may contain a single regime or a very large set of potential phases. The result is an endogenous theory of structural evolution and overlapping waves of technological development, based on explicit economic tradeoffs.

CHAOS

Radically simplified models of this kind generate nonlinear difference equations which are capable of generating deterministic, erratic behavior very much like the irregular fluctuations observed in reality. Moreover, as recent research has shown, these characteristics can be generic, i.e., present for more or less continuously varying classes of parameters, and ergodic, with long term frequencies of variable values converging to stable probabilities.

The nonlinearities responsible for these results are to be expected in other dynamic economic models. They occur because of the ubiquitous presence of nonnegativity restrictions on many economic variables, because of "natural" hypotheses such as liquidity traps, increasing and diminishing returns, and so on, and because of the quadratic nature of monetary values which always involve the multiplication of price and quantity.

In my opinion, it is not too early in the development of this theory to conclude that endogenously generated irregularity of these kinds is a very important ingredient in explaining the actual fluctuations of economic data. But this kind of chaos is a diversion from the main thread of the theory under consideration. There are much more crucial instabilities inherent in economic process than this one.

GLOBAL INSTABILITY AND INVIABILITY

These are suggested in extensive simulation expererience with empirical, recursive programming models. In general it has proven to be a nontrivial task to find parameter values that lead to convergence or even to viable solutions. Indeed, the typical model will work for a time, mimicking with more or less verisimilitude an actual history of some region or economic sector, but then becomes even less stable and stops working altogether. Models that stop working are called inviable. Their analogs in the real world are bankruptcies, banking system collapses, hyper-inflations and revolutionary economic breakdowns.

The latter forms of instability are relatively rare, but bankruptcy is a normal and continuing part of the working of an advanced economy. In this sense, inviability (global instability) is a further characteristic of complicated dynamics of individuals well captured by the theory and models we have put forward. But that poses a problem: if economies are inherently inviable, what keeps them running?

Viability is organic: life-bearing systems are not maintained for individual components, which are, for individuals, globally unstable subsystems that eventually disappear. Rather, the forces of change and development are acted out on a level that transcends individuals, a level within which the dynamics of reproduction, of birth and death, determine the viability of populations.

Societies adopt a similar solution when they provide for bankruptcy proceedings and new technologies, new preferences and new organizations. A quite analogous process also operates within the individual organization, with respect to rules of conduct that govern behavior within them. These rules of conduct are constantly judged by the economic forces of survival, accumulation, decumulation and demise. They are modified or replaced from time to time by innovative acts of planning and management. Indeed, human culture generally is a population of rules and regulations that originated in numerous acts of innovation and assimilation. These too are unstable. Many have disappeared. Of those existent, only a few are flourishing.

The recursive programming model of boundedly rational economizing with feedback takes on expanded meaning once we accept the view that economic systems are unstable, and globally so. Indeed, this approach represents economic changes as a counterpoint of adoption and abandonment of alternative ways of conducting economic activity and alternative objects of material form. Beyond this endogenous dynamics of development, it points to the inherent tendency toward breakdown that can only be overcome by more general evolutionary forces.

EVOLUTION AND CREATIVE MORPHOGENESIS

Biological evolution consists of the genetic processes of variations interacting with forces of individual behavior and environmental selection.

The economic evolution that is our subject here is, of course, imbued with the broader biological process, but in addition consists of a cognitive process of variation and selection interacting with the complex adapting system of individuals and organizations. These mental acts operate through an intricate, generative Gestalt in which the mind, processing whatever inputs it has, generates a new thought and creates a new sequence of acts that embodies that thought in some new form that was not there before.

This creative faculty must lie at the foundation of rational processes of thought, and hence of all of economics. Rational thought, after all, requires the comparison of alternatives according to well formed values, the perception of the limits of choice, i.e. the set of feasible alternatives and the selection of a candidate from this set that best satisfies preference. Especially when involving the possibilities for future action, this process involves "imaginative rehearsal" of possible scenarios of what might happen, sequences of imagined act and consequence that form conscious stories of what might be. To choose rationally is to compare stories, to select one and then to design a sequence of actions that will make those stories come true.

The imagination is also required to imitate what someone else has already figured out what to do for himself, as in the enjoyment of a new piece of music, the adoption of a new way of allocating resources, a new technology, a new product for consumption. Even obeying an authority requires an imaginative rehearsal that can lead to actions never taken before. Thus, imagination is an intimate part of the exercise of both rational and nonrational thought and, to the extent that people make conscious choices in their daily lives, it is routine: we can say that every human possesses it to some, however limited, degree.

This faculty of imagination which plays its routine role in everyday life rises to an exalted position in the functioning economy when it leads to invention and innovation of new ways of doing things, new things to produce and consume, new rules of conduct, new forms of information, decision and organization and new understandings of ongoing physical and biological processes in the nonhuman world. For these are the elements of variation that feed the process of selection and evolution that keep the economy as a whole working in the face of individual bankruptcies and the breakdown of various institutional systems of action. Individuals who possess these capacities to a high degree are called entrepreneurs in the world of business. It is their particular role to fashion into being the mechanisms that allow an economy to work when its agents are boundedly rational, its transactions imperfectly coordinated, and its long-run behavior intrinsically and globally unstable.

Entrepreneurs are both the result of - and the mediator of - evolution, both in its narrow biological sense and in its broader cultural sense. Once a part of human culture, their activity does not switch on and off according to well-defined accounting messages or in response to carefully anticipated need. It functions more or less continuously, thereby providing a continuous source of perturbation to the analytical structures that define routine production, consumption and managerial activity. The implication is that economies will evolve whether they need to or not.

Thus, the very faculty that makes economizing modes possible in general, and which plays an essential role especially in rational planning, is the source of a continual flow of perturbations that would disrupt any equilibrium that should accidentally occur.

TRANSACTIONS AND MARKETS

Among the activities engaged in by individuals in the course of allocating resources are transactions. Transactions among agents are mutually interrelated actions involving the exchange of information and goods, and the establishment or modification and constraints on further action. Such behavior involves further aspects of disequilibrium and instability that have not been accounted for so far. Transactions occur in several different manners which involve traditional mores for collection and redistribution in primitive economies, more or less bureaucratically administered rules within the complex organizations of modern economies, and decentralized market processes among individuals and organizations. The latter have - in traditional economic theory - been represented by bargaining between individuals in isolation or in a sequence of bilateral negotiations among freely associating traders, or as structured auctions, bidding systems, or negotiational procedures. These Bidding-Negotiation-Bargaining forms, which describe bilateral trade among nations, real estate transactions and the formulation of wage contracts, are of considerable importance, but like other aspects of rational activity are extremely time consuming and resource intensive. Though characteristic of market economies in early stages of development, they are increasingly supplanted as development proceeds by two fundamentally different processes of exchange.

Most evident on the retail scene are stores, which are nothing more nor less than inventories on display, as noted above. Almost as evident, and perhaps even more important, are Order-Delivery-Information systems that govern most wholesale, construction and heavy investment transactions. Individuals and organizations order goods. Producers, warehouses and stores receive orders and either fill them or delay delivery, adjusting their order backlog accordingly. Even the stock market, which is often thought of as an example of a competitive market, works in part on the basis of order-delivery-information systems with special broker-specialist agents.

From a physical point of view, these latter two Inventory-Order-Price Adjustment market types are stock-flow mechanisms that mediate transactions among agents using periodic price-adjustment rules. No doubt the specific character of the commodities involved, such as their storability, their time period of production or their relative cost, influence or determine what type of market mechanism is used in transactions involving them. But a noteworthy fact is that Bargaining-Negotiation-Bidding processes are not pervasive in the real world. Indeed, one could imagine an economy where they were absent altogether and exchange occurred using Inventory-Order-Price-Adjustment procedures exclusively. The basic virtue of these latter mechanisms is that they enable exchange to take place when supply and demand are not equated at prevailing prices. The participants need not postpone other activity while a sometimes interminable process of haggling works itself out.

When studied in highly simplified, experimental settings, direct exchange systems based on Bidding-Negotiation-Bargaining sometimes converge rapidly to competitive equilibria. These settings may be typical of some markets that are held at a single place at periodic intervals for relatively short periods of time, with relatively small numbers of people, such as auctions. Other markets, however, lack these characteristics. They are held continuously, can involve large numbers of individuals whose participation is not simultaneous, but strung out over time, and who may be separated by great distances. Inventory-Order-Price Adjustment mechanisms make such markets possible.

Certainly, markets of the latter type did not always exist. Their creation, however, introduced new avenues for exchange and with them, new possibilities for specialization in production, while enabling all this to happen in a decentralized, imperfectly coordinated flow of disequilibrium action. They have played, therefore, a crucial role in the progress of technological development and the growth of income and wealth. They provide a good example of how entrepreneurial activity has led to an evolution in the form and number of economic institutions.

MARKET INSTABILITY, UNCERTAINTY AND EXPOSURE

At the same time their complexities of dynamic interaction enhanced the conditions for disequilibrium, complicated change and inviabilities. The data of modern financial and commodity markets reflect this.

Markets both create opportunity and introduce exposure. They widen the scope of choice; they also expose participants to a widened range of uncertainties about the values of stocks and flows, of goods in exchange and even of access to the market system itself. Because of these uncertainties, and the realization from time to time of inviabilities due to exposures to an unpredictable fluctuation in values, some individuals and organizations are made worse-off by the system.

A further complicating force in decentralized exchange is the fact that individuals are routinely exposed to asymmetries in the power of bargaining, in part (and fundamentally) because individuals vary in their cognitive capacities, and because they vary in the initial conditions they bring to every act of exchange. These lead to asymmetries in the costs and benefits of exchange, and of changes over time in these costs and benefits. Such assymmetries lead to dissatisfaction not just with exchange but with the system of exchange.

POLITY

Implicit in the market system is a cooperative agreement to engage in peaceful, voluntary exchange on terms specified by the system. When plans cannot be realized and contingent tactics fail, individuals face the catastrophic risk of imminent demise. At such time, the prevailing system comes under review. Furthermore, consider asymmetries in bargaining power which, when extreme enough, can cause a breakdown in the system of voluntary exchange, and usher in a system based on coercion or deceit. Thus, market exposure creates constituencies for organizational innovation and motivates that enticing alternative to voluntary exchange called plunder: the taking by force or deception of what is possessed by another.

To prevent the destructive tit-for-tat of plunder or the fury of revolutionary breakdown in the prevailing economy, the participants of an economic system must develop a generative process of polity that allows for changes in the rules of economic conduct, an avenue for politicoeconomic morphogenesis so that recourse is restored for those who stand to lose too much or too often. If an equilibrating economy could be established, such a polity could (in the absence of creative thinking) wither away, leaving a fixed system of institutions and rules, or merely a collection of individuals with no institutions or rules at all. It would verge ever closer to a competitive or communal ideal in which everyone planned to do their best, everyone carried out their plan and no organization could be put forward to which anyone would object. But this cannot be the ideal world which can actually be brought about. Because individuals have limited powers of cognition and communication, and because of complicated dynamics and non-uniformity of preferences and goals, equilibrium behavior cannot emerge. It must, therefore, always be the case that disequilibrium persists and with it the potentially catastrophic exposure and the asymmetries in the costs and benefits of participation.

To eliminate disequilibrium is to destroy the system of decentralized, discretionary action and to replace it with one of administered rules of behavior based on tradition, imitation and obedience to authority. It is to constrain rationality and limit creative morphogenesis to operate within the bounds of established bureaucracy.

Alternatively, a society can embrace disequilibrium in a dynamic form of organization based on an alternative principle to that of a social equilibrium. That principle has been called Willing Participation. In such a society, most of the people most of the time will accept its working, will contribute to its functioning, will refrain from plunder and will defend it from any force from inside or out that would attempt to supplant discretion by coercion. Rationality, in such a system, cannot operate exclusively according to exact laws of deterministic, dynamical systems, because the functioning of these laws is too complicated; because their intrinsic working leads to unpredictable change and again and again to inviability; and because the more creative participants in the economy continually perturb it with wholly unanticipated possibilities for change.

Instead, such a system of polity must rest on access to its instruments, and because people will not generally agree, it must allow access to instruments of argument, persuasion and debate; modes of mental conduct and communication that go beyond the economizing modes of behavior and form the ingredients of democratic discourse and the basis for willing self-transformation. It becomes the medium through which institutions evolve. Subject to the opportunities and limitations of political process - and these limitations are severe - the people in such a system possess a freedom limited by the rules and operation of their collectively imposed and individually accepted system. Their potential participation in the continual evaluation of its components and in the process by which those components may be modified, replaced or augmented, is their exercise of liberty. It is not unlimited freedom, but a limited potential. It is the basis of their willing participation in a system whose functioning they sometimes regret.

The laws of such a system are not analytical in the Newtonian sense that governed the development of my argument through the concept of global instability. Instead, they are dialectical in the Aristotelian sense that understanding emerges from the free interplay of ideas and of the discussions about them. At the social level argument, persuasion and debate play the role that search and experimentation play in the faculty of mind that underlies individual volition. Their function is to synthesize - from the conflicting views of its boundedly rational participants - changes in the system of economy, of its distribution of wealth, of its rules and regulations that govern individual opportunity, of the understanding of how it works, and even of the values that guide rational thought. Thus the system, the new system so changed, will work - at least for a time - with minimal economic plunder and without debilitating social discord. The system of polity that provides the framework for their exercise coexists and codevelops with a market system. It both makes continuing disequilibrium possible, and mediates the evolution of mechanisms that perpetuate the process as a whole. I call it "the Dialectical Republic".

ACKNOWLEDGEMENTS

This paper draws on several recent expositions of mine, namely "Disequilibrium Economic Dynamics: A Post Schumpeterian Contribution", chapter 3 in R. Day and G. Eliasson (eds.), The Dynamics of Market Economies, North-Holland Publishing Co., 1986; and "The Evolving Economy", European Journal of Operations Research (forthcoming). It represents a further development of a line of thought that began with my first publication in economics, many years ago. As time has passed these ideas have gained in precision, generality and explanatory power. Discussions such as the present one have provided opportunities to reconsider and to improve their articulation. I am, therefore, grateful to the editors both for tolerating the incorporation of earlier material and for insisting that it be given a new package. For the relationship of the present study to other ideas, see the above references and the citations to the literature that they contain.

CHAPTER 4

The Balanced Path of Economic Development: A Fable for Growth Merchants

D. F. BATTEN

Once upon a time the tiny Kingdom of Entropia was enthralled by a great debate. 'This is a growing economy, but it still needs some qualitative improvements', some were heard to argue. 'Steady growth is best', came the reply, 'since we cannot live happily with sudden fluctuations or unexpected structural change'.

A few bemused onlookers called the debate gamesmanship. But most agreed that it was healthy if it would lead to a better understanding of how the tiny Kingdom might evolve in the longer term. So the King extended an invitation to all the leading scholars to gather together in one year's time in order to discuss the Kingdom's preferred path of economic development. In the meantime, he appointed a Committee of Enquiry to ascertain the basic facts about Entropian life.

The King's Committee soon reported that the population in Entropia was growing in accord with the familiar Verhulst equation of logistic growth. Being constrained by the Kingdom's limited resources and a ban on migration, the rate of population, P, was found to obey the law:

$$\dot{P} = \alpha P(N-P) - \gamma P, \qquad \alpha,\gamma > 0 \tag{1}$$

with α and γ pertaining to the birth and death rates of the population, and N defining the largest population which Entropia could currently sustain. The term (N-P) implied a saturation of the Kingdom's population to a finite steady-state level (Fig. 1), P°, where

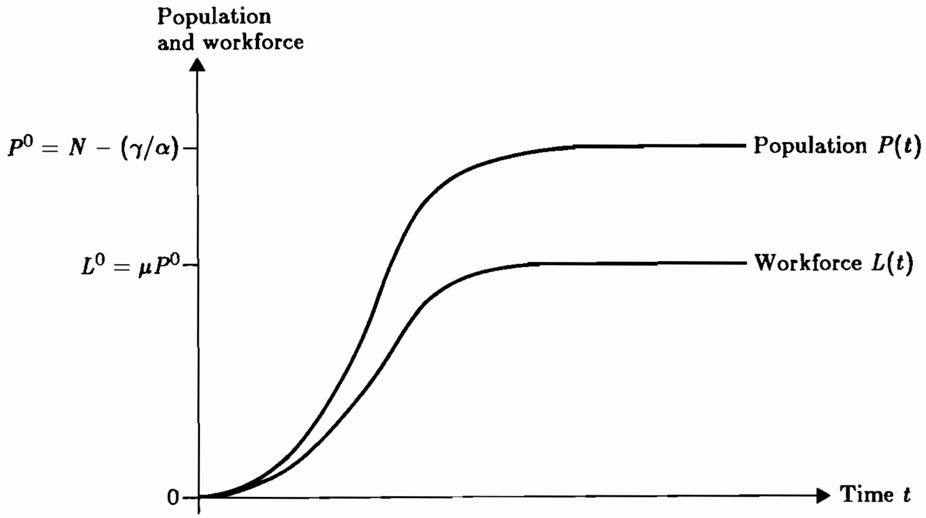


Figure 1 Growth of Entropia's Population and Workforce

$$P^{\circ} = N - (\gamma/\alpha). \quad (2)$$

The Committee's mathematician proudly announced that it was a simple task to demonstrate that this state is asymptotically stable whenever it exists (i.e. $P^{\circ} > 0$), whereas the trivial state, $P^{\circ} = 0$ – corresponding to the extinction of Entropia's population – is unstable!

For a moment, the King appeared satisfied with this result. Then his demographic advisor furtively pointed out that Entropia's population was currently within five percent of this stationary state. A nod of agreement from the Committee's economist signalled that, since the number of working Entropians was a fixed proportion, μ, of the population, the work force was rapidly approaching a similar capacity constraint. Suddenly, the stark truth dawned upon the King. The size of his Kingdom was finite.

Turning quickly to the Committee's economist the King asked: 'Before I am obliged to reconsider our migration policy, what is the kingdom's state of economic growth?' To this question came the confident reply that Entropia was a competitive economy making full and efficient use of its scarce factors of production, namely labour and capital, to produce a single, all purpose commodity. It was estimated that the efficiency of Entropian capital was increasing at a steady rate λ, and that the productivity of the labour force had hitherto been advancing at a similar rate.

Although uncertainty prevailed concerning the economy's expansive capabilities in the very long run, Entropia's production capacity, $x(t)$, at time t was assumed to be the following function of available capital, K, and labour, L:

$$x(t) = e^{\lambda t}[K(t), L(t)], \quad \lambda > 0 \tag{3}$$

where $L(t) = \mu P(t)$. The economist's report acknowledged, however, that a detailed investigation of the state of Entropia's technology would be desirable following wide variations in the reported productivity and profitability of each firm.

The Committee had also sought counsel on the issue of economic development in the long run from a prominent supply-side economist known as Jo, whose views on capitalism were held in very high regard. Jo's advice began: "The essential point to grasp is that in dealing with capitalism we are dealing with an evolutionary process [16, p.82] ..... The fundamental impulse that sets and keeps the capitalist engine in motion comes from the new consumers' goods, the new methods of production or transportation , the new markets, the new forms of industrial organization that capitalist enterprise creates" [16, p.83] .

Following such a profound statement, some puzzled members of the Committee asked if this engine could be driven by an increase in the population or from changes underway in Entropia's physical and social environments. 'The essence of economic evolution is none of these' replied Jo, 'since I am speaking of a process of qualitative change'. "The opening up of new markets, foreign or domestic, and the organizational development from the craft shop and factory to such concerns as U.S. Steel illustrate the same process of industrial mutation – if I may use the biological term – that incessantly revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one. This process of creative destruction is the essential fact about capitalism" [16, p.83] .

'Are you suggesting that economic development is an endogenous process?' queried one of the Committee members. "Yes", Jo replied: "By development, therefore, we shall understand only such changes in economic life as are not forced upon it from without but arise by its own initiative, from within..... Nor will the mere growth of population and wealth be designated here as a process of development. For it calls forth no qualitatively new phenomena" [15, p.63] .

At this point, the ramifications of Jo's counsel began to dawn upon the King. First, he realized that to approach economic development as a factor exogenous to the economic system, as implied by equation (3), would be erroneous. Second, since such development is a creative process that replaces old states with new ones, it was largely foreign to the classical dynamic theories of balanced and periodic business cycles. Such endogenous and discontinuous changes might not merely disturb the steady state currently prevailing in Entropia, but may instead displace it forever. Third, he suspected that the potential rate of development in Entropia might be dictated more by the competence and creativity of his subjects than by their total number.

The King commended the Committee for its informative report. He then announced the provision of a generous prize for the discovery of Entropia's optimal development policy, emphasizing that the main objective of such a policy would be to determine those conditions under which innovative technological progess could evolve at a harmonious and sustainable pace.

And so it came to pass that all the leading scholars from within and beyond the tiny Kingdom duly assembled at the palace. Invited to speak first because of his pioneering scientific work on processes of self-organization, Ilio likened the competitive process of industrial innovation to the onset of structural fluctuations observed in population biology. 'The key to this competitive type of evolution is the appearance in small quantities of new species which we call <u>mutants</u> or <u>error copies</u>', said Ilio. 'One might imagine that by treating industrial innovations as fluctuations relating to the structural stability of the equations of industrial dynamics, a criterion might be developed concerning the long-term evolutionary prospects for Entropia's economy'.

To illustrate Ilio's suggestion, an economist called Davo offered the following analogy: 'In the absence of qualitative changes, we know that Entropia's rate of industrial development is limited by the availability of our scarce factors of production. Thus we can assume that industrial output grows in a similar manner to the population, i.e. as defined by equation (1). In such an equation, let us replace P by x_1, the current supply capacity of Entropia's industry, and relate α to the rate of entry of new firms into this industry and γ to the rate of exit of old firms.

Now suppose that some of the new firms commencing production at one stage are technologically superior. Instead of just copying our current "best-practice" technology, these firms introduce a new technology of their own to produce a better all-purpose product. Let x_2 denote the productive capacity of firms of this new innovative type and suppose that, after some time, their number is sufficiently large to enable their growth to be described by an equation of the following form:

$$\dot{x}_2 = \alpha_2 x_2(D - x_1 - x_2) - \gamma_2 x_2. \tag{4}$$

In equation (4), the values of α_2 and γ_2 are related to the entry and exit rates of the innovative firms, respectively. Furthermore, if the parameter D defines the total demand at time t for Entropia's all-purpose commodity, then the trajectory described by equation (4) implies that the market share of the new innovative firms depends partly on the difference between their entry and exit rates. This is in accordance with our recorded observations of competititve behaviour.

The rate equation describing the growth of those old firms already established in the industry takes an identical form:

$$\dot{x}_1 = \alpha_1 x_1(D - x_2 - x_1) - \gamma_1 x_1. \tag{5}$$

The state of the industry immediately before the innovative firms appear is given by

$$x_1(0) = D - (\gamma_1/\alpha_1) \text{ and } x_2(0) = 0.$$

If we rewrite equation (4) in the functional form

$$\dot{x}_2 = F(x_1, x_2) \tag{6}$$

our criterion for evolving away from $x_2 \approx 0$ reduces to the following simple requirement[1]:

$$\partial F/\partial x_2 > 0.$$

This means that the very presence of one or two innovative firms enhances the rate of entry of others, which may be called an autocatalytic process. Applying this criterion to equation (4) we find that if

$$D_2 - (\gamma_2/\alpha_2) > D_1 - (\gamma_1/\alpha_1) \qquad (7)$$

then the group of innovative firms steadily grows to occupy a niche in the industry (see Figure 2). D_1 denotes the total demand before, and D_2 the demand after the innovation.

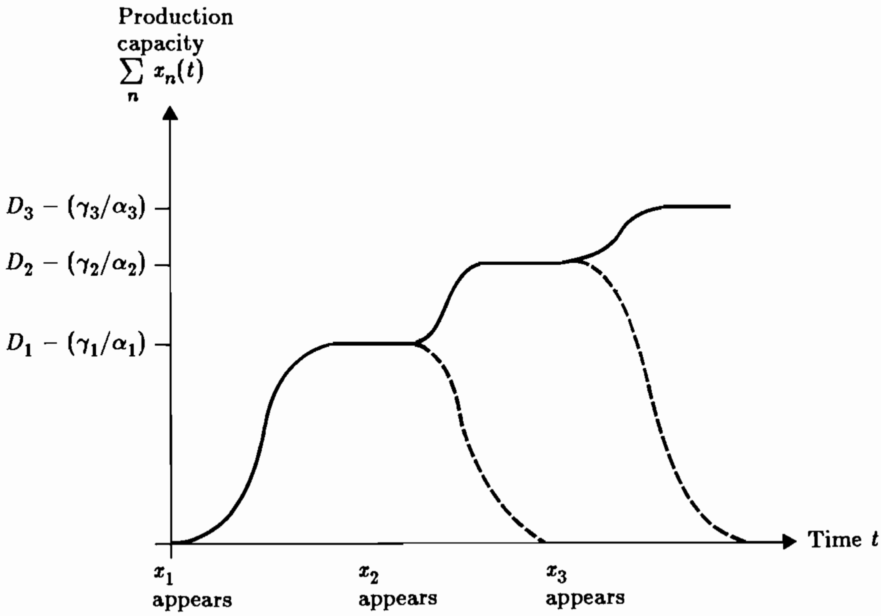


Figure 2 Productive Capacity Expansion of Entropia's Industry (adapted from [3])

The King thanked Ilio and Davo for their valuable insights, then invited his favourite economic advisor to respond to these ideas. Being an exceptionally creative economist in his own right, Ako offered the following comment: 'I am extremely impressed by this theory of self-organizing, dissipative systems. It seems that a new scientific paradigm has emerged. But perhaps I may be permitted to suggest that the notion of innovation implied in your theory is not the traditional economist's view; in which technological development is supposed to be influenced by additional state variables such as prices and profits. If I understand your model correctly, innovation is being portrayed as a structural change the timing of which is essentially uncertain. Those economists among us would doubtless wish to probe beneath this mathematical elegance in order to identify the specific catalyst.'

'My sentiments exactly', quipped the King in a rare moment of solidarity. 'But can you enlighten us, Ako, about the circumstances under which such technological progress might take place?'

'I would propose that the creation of new knowledge is the key catalyst. Entropia's production function (equation 3) ought to be modified in the following way:

$$x(t) = F[G(t),K(t),L(t)] \tag{8}$$

so as to consider the current capacity of the learning base, G(t), as an endogenous factor input. Then we could begin to explore how knowledge formation interacts with labour competence and capital formation in the Entropian economy.'

Upon hearing this statement, an elderly observer (who had hitherto listened in silence to the proceedings) rose to his feet and made the very simple heuristic suggestion that the whole Entropian economy, and its numerous agents, operate in a manner akin to learning systems, being basically governed by Volterra-Lotka equations. Introducing himself as Cesaro, he argued eloquently that the concept of a learning society offers a very powerful route to a unified theory for genetic evolution, ecology, sociology, and economics. Some Entropians murmured that the plethora of logistic market penetration models amply describing the long-term behaviour of so many different foreign markets certainly added strength to his argument[2]. But a stubborn group of economists were not enthusiastic about this abstract generalization. To steer the discussion back on a more familiar course, they nominated a scholar named Katso to speak on their behalf.

Katso spoke. 'I begin with a remark concerning the learning process. In the Entropian world of dynamic competition, each firm is constantly striving for a better production method. As Ako has told you, a firm may succeed in putting a new production technology into practice through its own promulgation of new ideas. We may view this as learning from within, and call the process innovation. Another firm could instead learn only from observing the techniques of other innovative firms. Let us refer to this process of adopting one of the methods currently in use by other firms as imitation. "It is the dynamic interaction between the continuous and equilibrating force of imitation and the discontinuous and disequilibrating force of innovation which governs the evolution of Entropian industry's state of technology" [8, p.187] .

'Bravo', cried an enthusiastic group of Entropians. 'Your realistic theory suggests that Entropia's industry will not reach a neoclassical equilibrium with perfect technological knowledge even in the long run. We have observed that the actual production method of many of our firms always lags behind the best-practice technology. Can you tell us more about our production techniques?'

'Entropian industry consists of a large number of firms competing with each other to produce one single, all-purpose product. Just as Davo has suggested, the industry as a whole is continuously experiencing a turnover of firms as new ones enter and old ones withdraw. Since there is no automatic awareness of the best-practice technology, each firm is most likely using a slightly different production technique. Furthermore, each chosen technique embodies the technology (and hence capital equipment) of its date of construction. Observe that since the market price of your all-purpose product is identical everywhere, the smaller each firm's production cost per unit of output (c_i), the more profitable will be the corresponding production technique. We shall refer to the unit profit of production method i as π_i.

'Since a limited number (m) of production techniques prevail in Entropian industry, we can arrange them in order of decreasing best-practice, thus:

BEST PRACTICE $\quad c_m < c_{m-1} < \ldots\ldots < c_i < \ldots\ldots < c_1 \quad$ WORST PRACTICE

We can then describe the overall state of technology in Entropia at a given point in time as the distribution of these different production techniques (or unit production costs) across all our firms. If we let $f_i(t)$ stand for the relative market share of firms whose unit cost equals c_i (and whose unit profit equals π_i) at time t, and $F_i(t)$ denote the cumulative market share of those same firms at the same time, we can depict our entire Kingdom's state of technology in the form shown in Figure 3. We must remember, however, that the profile of this distribution will be constantly changing, as new firms enter and old firms update their technology or withdraw.'

'What conditions, then, must be met in order to guarantee technological progress?' asked the King.

'Each firm, whether new or old, can attempt to put a new production technique into practice (innovate) or it can copy another technique already in operation (imitate). If the above concepts are clear and my assumptions are granted, I wish to introduce the following proposition regarding imitation'.

'A proposition, a proposition', the assembled crowd shouted. It was evident that Entropians were enchanted by the prospect.

Katso resumed. ' Proposition: Let us assume that the probability of any firm imitating an existing production technique of unit cost c_i is proportional to the market share, $f_i(t)$, of those firms which already employ that technique at time t. The unit cost of the technique which the firm implements will be lower than the one currently in use but at least as high as that of the best-practice technology. Then, since the total number of firms in Entropian industry is large, we can approximate the change in its technological structure by the following series of differential equations:

$$\dot{F}_i(t) = \beta F_i(t)[1 - F_i(t)] , \qquad i=1,m \tag{9}$$

where $\dot{F}_i(t)$ denotes the time derivative of $F_i(t)$ and β is an imitation coefficient which reflects the speed of the spread of information between firms and the amount of information needed to make a positive imitation decision.'

It took only a moment's reflection on the part of some of the crowd to recognize that each of the above differential equations describes a similar logistic path to the growth process depicted in Figure 1. Taken together, the envelope of S-shaped curves postulated by Katso portrays a bandwagon phenomenon (in terms of imitation) which would eventually swamp the whole of Entropia with best-practice technology. This

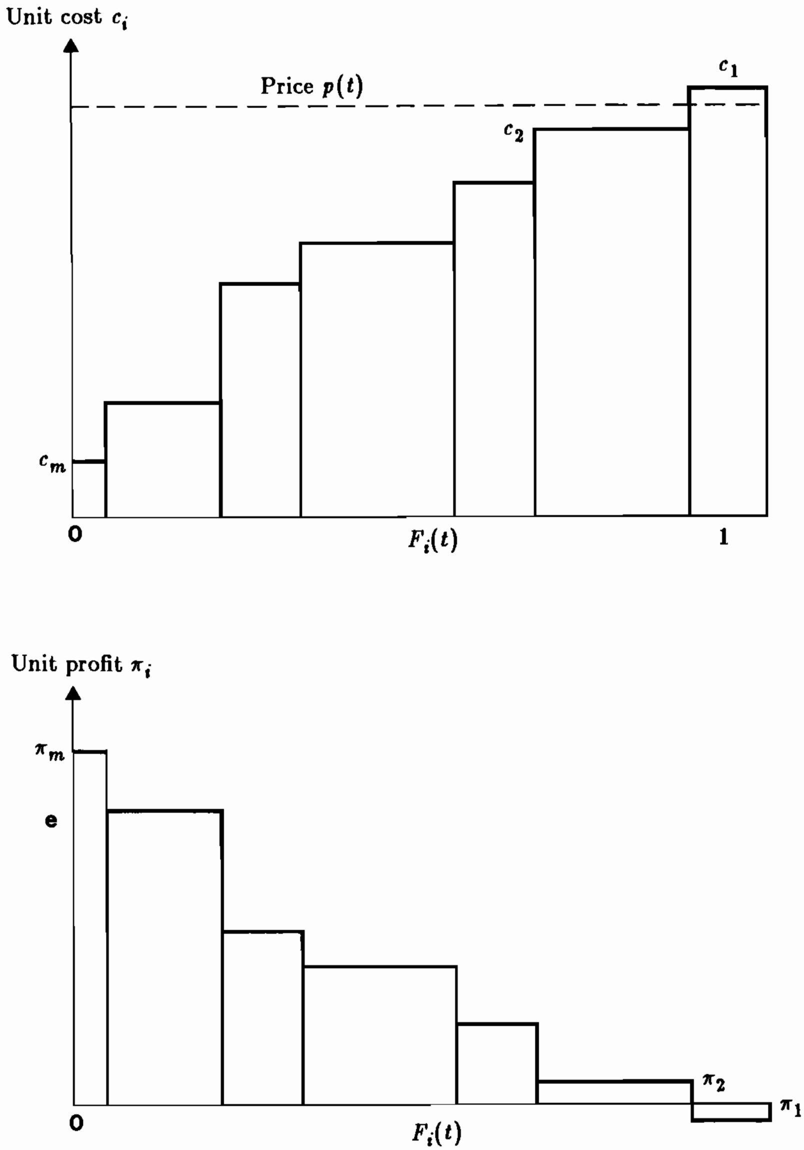


Figure 3 State of Technology and Profitability of Entropian Firms (adapted from [8])

process of cumulative diffusion is illustrated in Figure 4.

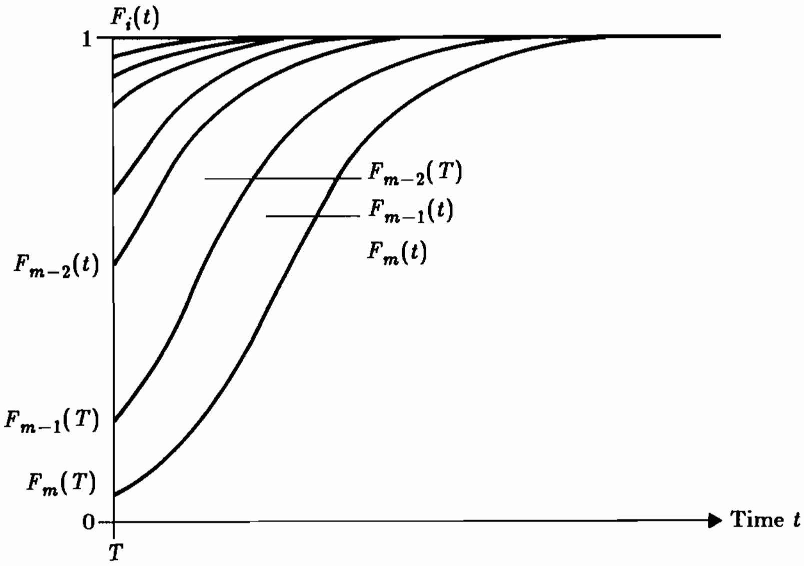


Figure 4 The Evolution of Entropia's State of Technology through the Cumulative Process of Imitation (adapted from [8])

Katso continued: 'The limiting state, in which all Entropian firms have adopted the best-practice technology, corresponds to the familiar paradigm of neoclassical economics. However, this very tendency towards technological uniformity is in itself a strong catalyst for the introduction of a new and better production technique. "Indeed, to destroy the stalemate brought about by the imitation process and to create a new industrial structure is the role our capitalist economy has assigned to Schumpeterian entrepreneurs or to innovative firms" [8, p.169] .

Katso had now completed his elegant discourse concerning the envelope of imitation curves, and was modestly preparing to proceed with a formal discourse on innovation. Anticipating that this might rest on the very reasonable assumption that someone succeeded in implementing a new production technique (whose unit cost, c_{m+1}, is less than c_m), the King chose to intercede: 'Thank you, Katso, for your interesting proposition. It would be instructive to analyse how a new wave of imitation is spawned by the introduction of a new technology. Nonetheless, I prefer to discover what triggers such innovative progress in the first place. Technological uniformity may in fact be one catalyst for innovative progress, but such a condition would then confound your basic proposition. I wonder if the profitability criterion is not a sufficient incentive for innovative firms with a new production technique to enter the market, and for Entropian output to suddenly expand at a much greater rate than by imitation alone?'

At this juncture, a clever economist named Bojo, who had been waiting restlessly throughout Katso's presentation, stood up and said: 'An important feature of Katso's proposition is that it describes the evolutionary path of the cumulative proportion of Entropian firms employing each production technique, but it does not define the actual production level of each. A similar problem arises with respect to market penetration models which portray the evolution of each competitor's market share. The profile of market share development over time tends to exhibit a reasonably smooth regularity which is largely unaffected by differences among individual firms. Rather than abstract from supply-demand mechanisms, it would be preferable to introduce a framework which addresses such issues, and in which the parameters of any substitution process may function as a reflection of changes in relative prices in combination with changes in taste and technology.[3]

Bojo continued: 'Suppose that an entirely new all-purpose product were to be introduced in Entropia, and that this product possessed certain characteristic advantages compared with your current product. This superiority might prove sufficient to expand the market in terms of total demand (as depicted in Figure 2), with the potential market share for the new product depending on its qualitative characteristics as well as its price relative to that of your current product. Given that a new equilibrium may be attainable, it is possible that the process of firm-by-firm imitation postulated by Katso will generate the following logistic pattern of growth in aggregate demand for this new product:

$$\dot{y}(t) = \beta' y(t)[D(p) - y(t)] \text{ , and } y(o) = D_0 \tag{10}$$

where y(t) is the level of demand for the new product at time t, β' is an imitation coefficient, D_0 is the equilibrium level of market demand for your current product, and D(p) is the new equilibrium level of market demand. For convenience, we shall assume a strictly linear downward-sloping demand-price relationship:

$$D(p) = a_0 - a_1 p(t). \tag{11}$$

We shall simplify the comparative characteristics of this new product by combining them into the composite form of a lower "qualitative price", resulting in a potential demand increase of $D(p) - D_0$. Thus equation (11) portrays the equilibrium demand curve for the new product, with the disequilibrium adjustment process being determined by equation (10). This latter process corresponds to the envelope of diffusion curves described earlier by Katso (see Figure 4.)'

'But how is this price, p(t), to be determined?' asked one of the King's advisors. 'Surely it cannot be an exogenous factor in this process of adjustment?' another queried. Bojo resumed the argument.

'To answer such questions, we must turn from the demand side and explore the growth of your capacity to supply the new product. Our international studies reveal that the rate of increase in capacity depends on the profitability of producing the new product. Let the profit per unit of new output, $\pi(t)$, be defined as

$$\pi(t) = p(t) - c(t) \tag{12}$$

where the unit cost, c(t), includes the cost of all the pertinent factor inputs: labour, materials, capital and knowledge. Let γ denote the fraction of profit ploughed back into capacity expansion, and let ρ be the ratio of external to internal funds invested at any time. It then follows that the growth rate of capacity, $\dot{x}(t)$, will be linked to the rate of profits by

$$\dot{x}(t)/x(t) = \gamma(1+\rho)\pi(t). \tag{13}$$

Combining (12) and (13), we have

$$\dot{x}(t) = \alpha x(t)[p(t) - c(t)] \tag{14}$$

where $\alpha = \gamma\ (1+p)$. [4]

'I have the impression that a crucial point in your exposition may have just been reached', interrupted Katso. 'It is conventional to assume that the continuous inflow of technological knowledge and process refinement serves to constantly reduce the unit cost of any new product, so that we usually have

$$\dot{c}(t) < 0.$$

This downword-sloping learning curve is often presumed to take an exponential form. In Entropia's case, however, we know that inelasticities exist in the supply of labour, implying that as the adoption of the new innovation grows and output expands, a point may be reached when the composite unit cost of factor inputs (and, in particular, wage costs) will start to rise rather than fall. Let us now assume that we have reached that point where the unit cost of composite inputs, c(t), increases (for simplicity, we assume a linear relationship):

$$\dot{c}(t) = c_0 + c_1 x(t). \tag{15}$$

Such increases in unit cost can serve to bring Entropia's economy back into equilibrium.'

At this point, Davo took up the argument: 'Combining (14) and (15), we have

$$\dot{x}(t) = \alpha\, x(t)[p(t) - c_0 - c_1 x(t)] \ . \tag{16}$$

Given that technology (c_0, c_1) is fixed, the growth of Entropian production will reach a ceiling when the unit cost is driven as high as the market price. This is all in accord with basic product cycle theory.

Referring to (10) and (16), we can see immediately that both capacity growth and demand growth are governed by logistic processes in which the price of the new product plays a key role. In our closed Entropian economy, the growth rates of demand and capacity cannot remain out of step with each other for very long. Price will serve as an equilibrating factor to balance the two growth processes (in the absence of further innovations). We may refer to such a transition as a balanced adoption path.

Along this balanced path, we have

$$\dot{x}(t)/x(t) = \dot{y}(t)/y(t) \text{ and } x(t) = y(t). \tag{17}$$

By substituting (16) into (11) and the result into (10), we reach the fundamental discovery that

$$\dot{x}(t) = \{\alpha/(\mu'a_1+\alpha)\}\ x(t)[(\mu'a_0-c_0)-(\mu'+c_1)x(t)] \tag{18}$$

or, in simpler terms, that

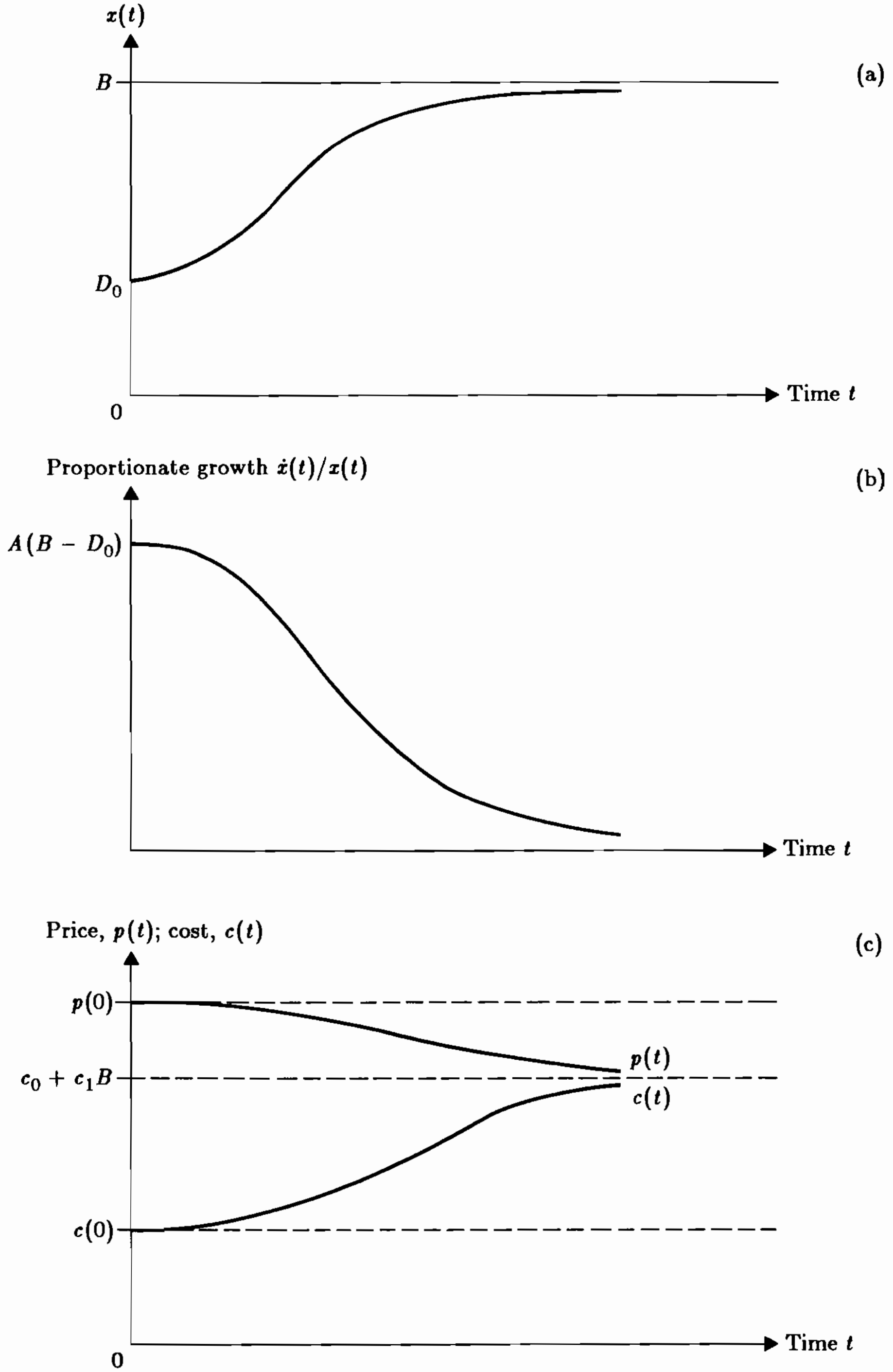
$$\dot{x}(t) = Ax(t)[B - x(t)]$$
$$x(0) = D_0$$

where $A = \alpha(\mu'+c_1)/(\mu'a_1+\alpha)$ and $B = (\mu'a_0-c_0)/(\mu'+c_1)$.

Entropia's balanced path of economic development is therefore a logistic process whose key parameters reflect the joint dynamics of demand growth and capacity growth.[5] The saturation level of output, B, depends on the rate of imitation, the equilibrium level of demand, and the factur input schedules defined in equation (15). The balanced adoption coefficient, A, depends on two rate constants (which govern the dynamics of demand growth and capacity growth) and two elasticities (one of demand and the other of supply). Note that the balancing factor is given by the product AB, and that this is independent of the elasticity of supply.

Figure 5 illustrates the main implications which may be drawn from our explanation of Entropia's balanced path of development. The introduction of a new, all-purpose product acts as an economic catalyst which creates a potential for adjustment up to B, with the scope for new economic growth being $B - D_0$. We may fruitfully capitalize on this innovation by encouraging Entropian firms to switch progressively from their old product to the new one in such a way that the demand for the innovation and the capacity to produce it grow at a harmonious rate.

Figure 5 also shows the effect of the balanced development process upon production capacity, price, unit cost, and profitability. As the new industry expands, the increasing cost of the composite input pushes the unit cost along the path described by $c(0) - c(t)$, tending toward the stationary level $c_0 + c_1 B$. The price of the new



<u>Figure 5</u> Entropia's Balanced Path of Innovation Diffusion (adapted from [10])

product will also tend toward the same stationary level so that, over time, profit margins are progressively reduced. As the price of the new product falls, for adopters there is a potential profit to be gained from switching to the new product.[6] Thus profitability will be viewed differently by the producers of the innovation and those who might imitate it; since it varies systematically over time for both, but in opposite directions'.

At this juncture, Jo was seen to nod his approval of this dual view of profit as the temporary reward for the innovator's entrepreneurial activity; being destined for destruction as the more competitive diffusion process proceeds. So much so that Jo rose to his feet and declared: "Profit..... attaches to the creation of new things, to the realization of the new value system. It is at the same time the child and the victim of development' (16, pp. 153-154).

Realizing that an important consensus may be near, the King asked Bojo and Davo to summarize their findings. Davo concluded: 'Given the fundamental data available to us, Entropia's balanced path of innovation diffusion may be governed by just two factors: the adjustment gap $(B - D_0)$ and the dynamic attributes of demand and capacity growth, as summarized in the balanced adoption coefficient A. It is to the control of these two factors that further attention must be directed.

The Entropians were clearly impressed by all these suggestions. But they were a practical and inquisitive people who were soon full of queries. What must we do if we are not already on this balanced path of development? Can we abide by this rule even when Entropia's economy is out of equilibrium? How do we ascertain the best time for innovation rather than imitation?

Before any of the assembled scholars could respond to these questions, the King interceded. 'Please restrain yourselves for a moment. It is time for an assessment of all these valuable ideas by my two invited judges, who will announce the final prizewinner.' He then introduced Rico and Sido, two prominent American economists who were widely known for their evolutionary theory of economic change.

Rico began. 'We have listened with much admiration to the theories and prognoses about the state of development in this tiny Kingdom. There is valuable wisdom in all that has been offered. Thus we found it extremely difficult to select one winner from such interesting contributions. Before announcing our decision, Sido and I would like to offer a few ideas of our own which may serve to clarify some of the uncertainties which remain with respect to Entropia's future.'

Sido continued: 'I ask you first to recall Ilio's notion of comparative fitness in population biology, since the concept has much in common with the relative profitability of different production techniques. Just as fitness clearly depends on the characteristics of the environment confronting the species, so profitability depends on the market price confronting firms with similar yet inequivalent production methods. The final price in turn depends on the various methods of all firms operating at any one time. In fact, the essential character of Schumpeterian competition hinges on the exact technological diversity which exists among Entropian firms. It is the general robustness (or fitness) of the cumulative distribution of profitability or productivity across all firms (as depicted in Figure 3) which best describes the state of technological development in Entropia at any point in time.

Now consider the model outlined by Davo and Bojo. If the initial price, p(0), and the corresponding output level D_0, are driven out of equilibrium by the introduction of an innovative product, a process of imitative adjustment will occur in order to reach a new equilibrium. In Figure 6, the lines d — d and s — s represent the adopters equilibrium demand curve and the producers' equilibrium supply curve, respectively; intersecting at the output level B. From the disequilibrium point $(p(0), D_0)$, price and output will follow the arrowed path until B is reached and the new product becomes part of a stationary economic routine. A similar adjustment process may be precipitated by a shift in either schedule. As Katso explained, it is this dynamic interplay between the equilibrating force of imitation and the disequilibrating force of innovation which governs the development of Entropia's economy in the long run. For this fundamental insight, we declare Katso the prizewinner.'

At this announcement, the King shouted rather excitedly: 'Surely we can now identify two warning signals which, when considered together, might help us to identify the time for qualitative change. First, we must monitor the shape of Entropia's technological profile over all firms, as depicted in Figure 3. If this is concave to the origin, then Entropian industry as a whole is quite robust. If it is

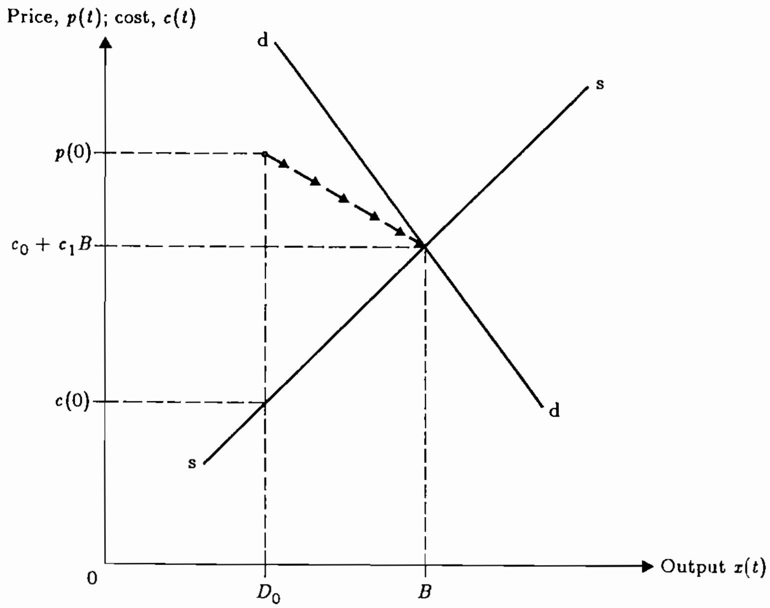


Figure 6 The Path of a New Equilibrium State in Entropia (adapted from [10])

linear or convex, however, then a significant group of firms will be vulnerable to small rises in our unit cost. At best, they must be willing to renew their technology; at worst, they must discontinue their productive activity.

Second, we must measure the degree of equilibrium prevailing at any point in time. As Entropia approaches a state of equilibrium and technological uniformity, we must encourage new entrepreneurs and new firms to generate new ideas (see Figure 7). Only by innovation and disequilibrium can Entropia overcome its scarcity of resources and find a new and better balanced path of economic development.'

The crowd finally dispersed, happy that a formula had been found for their Kingdom's future. Naturally, a few skeptics remained. Those who had argued for steady growth in Entropia at the outset of the great debate could not accept the desirability of a disequilibrium state. Others remained uncertain about the nature of desirable innovative activity and the legality of imitating an innovation.

To these consternations, the King responded in the following manner: 'We can never expect economic life in Entropia to be easily predictable or perfectly stable. Inevitably, there will be many uncertainties. Just remember the name of our tiny Kingdom! But if we are willing to breed or adopt innovations at the appropriate times, and not to cling too tightly to our old ways, then we may be assured of reaching a new

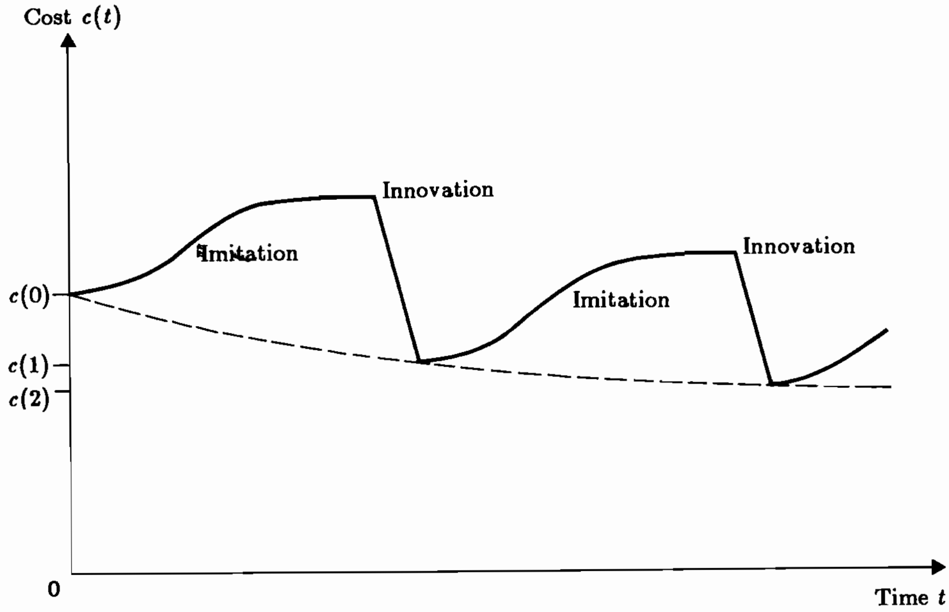


Figure 7 Successive Paths towards and away from an Equilibrium State in the Kingdom of Entropia

balanced path of economic development!

To further this end, the King proclaimed balanced development a national objective and instituted special incentives for timely innovations and rapid imitations. A series of balanced paths was duly attained and thus Entropians enjoyed, subject to the linearity assumptions needed to reach (18), a remarkably dynamic economy ever after!

NOTES

[1] The stability of the pair of equations (4) and (5) will be determined by a characteristic equation of degree 2. This equation has two roots, one of which has a value close to that of the original characteristic equation and, in particular, has a negative real part. The problem we are dealing with concerns the structural stability of the original system with respect to pertubations which increase the order of the

differential system. Allen [1] suggested that the criterion for evolving away from values of x_2 which are practically zero is that the determinant of the new characteristic equation should be such that at least one root contains a positive real part. Such a criterion does not depend explicitly on the properties of the original system alone, although a higher level of productivity will ensure that x_2 survives.

[2]The analyses conducted by Fisher and Pry [7] and Peterka [14] are two of the most interesting.

[3]For a detailed discussion of the Lancasterian foundations of this choice framework, see Batten and Johansson [4].

[4]At this point, Bojo was seen to wave gratefully to V. Peterka [14] for suggesting this formulation.

[5]As he reached his interesting conclusion, Davo hastened to declare that this unified framework came from J.S. Metcalfe [10].

[6]Note that p(t) need not necessarily fall (as depicted in Figure 5c) for the excess of p(t) over c(t) to contract.

ACKNOWLEDGMENTS

The author begs the indulgence of those excellent scholars whose ideas and research results have been discussed under the pennames designated below. He accepts full responsibility for any misunderstandings or misrepresentations of their work. In addition, the author is indebted to E.S. Phelps for his fascinating fable about the kingdom of Solovia [13], which certainly inspired this narrative.

BIBLIOGRAPHY AND REFERENCES

[1] Allen, P.M., Evolution, population dynamics and stability, Proceedings of the National Academy of Sciences, 73, 1976, pp. 665-668.
[2] Andersson, Å.E., Structural change and technological development, Regional Science and Urban Economics, 11, 1981, pp. 351-361.
[3] Batten, D.F., On the dynamics of industrial evolution, Regional Science and Urban Economics, 12, 1982, pp. 449-462.
[4] Batten, D.F. and Johansson, B., Product substitution and spatial cycles, forthcoming in A.E. Andersson et.al., eds., Dynamic Analysis of Spatial Development, North-Holland, Amsterdam, 1987.
[5] Businaro, U.L., Applying the biological evolution metaphor to technological innovation, Futures, 15, 1983, pp. 463-477.
[6] Eigen, M. and Schuster, P., The Hypercycle, Springer-Verlag, Berlin, 1979.
[7] Fisher, J.C. and Pry, R.H., A simple substitution model of technological change,

Technological Forecasting and Social Change, 3, 1971, pp. 75-88.
[8] Iwai, K., Schumpeterian dynamics: an evolutionary model of innovation and imitation. Journal of Economic Behaviour and Organization, 5, 1984, pp. 159-190.
[9] Marchetti, C., Society as a learning system: discovery, invention and innovation cycles revisited, Technological Forecasting and Social Change, 18, 1980, pp. 267-282.
10] Metcalfe. J.S., Impulse and diffusion in technical change, Futures, 13, 1981, pp. 347-359.
[11] Nelson, R.R. and Winter, S.G., An Evolutionary Theory of Economic Change, Harvard University Press, Cambridge, Ma., 1982.
[12] Nicolis, G. and Prigogine, I., Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations, John Wiley and Sons, New York, 1977.
[13] Phelps, E.S., Accumulation and the golden rule, in: A. Sen, ed., Growth Economics, Penguin Books, Middlesex, 1970.
[14] Peterka, V., Macrodynamics of technological change: market penetration by new technologies, Research Report RR-77-22, 1977, International Institute for Applied Systems, Laxenburg, Austria.
[15] Schumpeter, J.A., The Theory of Economic Development, Oxford University Press, New York, 1961, (first edition 1934).
[16] Schumpeter, J.A., Capitalism, Socialism and Democracy, Allen and Unwin, London, 1976 (first edition 1943).

LIST OF SCHOLARS:

Ako	- Å.E. Andersson
Bojo	- B. Johansson
Cesaro	- C. Marchetti
Ilio	- I. Prigogine
Jo	- J. Schumpeter
Katso	- K. Iwai
Rico	- R. Nelson
Sido	- S. Winter

Chapter 5

Competitive Growth of Firms in an Industry

E. AKIN

1. INTRODUCTION

In chapter 10 of their book "An Evolutionary Theory of Economic Change", Nelson and Winter (1982) describe a model for the dynamics of competitive growth among firms in an industry. As its title suggests, the book is generally inspired by evolutionary biology, and their model can be analyzed completely using mathematical techniques recently developed for applications in population genetics.

In a wide class of models one must impose unrealistic assumptions – such as equal numbers of factors and firms – to obtain isolated equilibria. Our analysis contains examples which illustrate the way in which the degeneracies introduced by relaxing such assumptions can be handled.

We demonstrate that every initial state tends to equilibrium in this model, subject to structural assumptions possibly vulnerable to criticism (see section 3). However, our principal conclusion – that enriched technical diversity can result in a decrease in total output – is both robust and surprising. We shall show that homogeneity assumptions can restrict attention to misleading special cases.

2. BEHAVIOR OF THE MODEL

Our industry consists of firms indexed by $j = \{1,...,n\}$; $x_j \geq 0$ is the output per unit time by firm j of the common industrial product. So at any moment the industrial state is described by a vector x, in the set of nonnegative vectors, R^n_+. The total output denoted $|x|$, is $\Sigma_j x_j$.

The firms employ factors indexed by $i = \{1,...,m\}$. $a_{ij} \geq 0$ is the demand at unit output by firm j for factor i. We assume fixed technology and constant returns to scale in production so that the a_{ij}'s are constants and the total industrial demand for factor i is $a_{ix} \equiv \Sigma_j a_{ij} x_j$.

By p and w_i we denote, respectively, the price of the product and the wage for factor i. So the unit profit for firm j is

$$\pi_j \equiv p - \Sigma_i w_i a_{ij} \tag{2.1}$$

and its net income is $x_j \pi_j$.

We introduce dynamics by assuming that each firm grows at an absolute rate proportional to its net income or, equivalently, at a relative rate proportional to its unit profit. Assuming a common proportionality constant we can rescale time or price to make it unity. We then obtain the following system of differential equations:

$$dx_j/dt = x_j \pi_j = x_j(p - \Sigma_i w_i a_{ij}). \tag{2.2}$$

Define the market share or relative size s_j of firm j by the ratio of its output to the total product, i.e. $s_j = x_j / |x|$. Thus, if $|x| > 0$ (some firm is operating), the distribution vector s of relative sizes satisfies: $1 \geqq s_j \geqq 0$ and $\Sigma_j s_j = 1$. Employing system (2.2) it is easy to verify that the dynamics of total output and relative size are given by

$$d|x| / dt = |x|\bar{\pi} \quad \text{with} \quad \bar{\pi} \equiv \Sigma_j s_j \pi_j = p - \Sigma_{i,j} w_i a_{ij} s_j \tag{2.3}$$

and

$$ds_j/dt = s_j (\pi_j - \bar{\pi}). \tag{2.4}$$

Thus, the growth rate of s_j is the amount by which the profitability of firm j exceeds the industry average where the average itself is weighted by relative size.

If the prices - and consequently the unit profitabilities - are assumed to be constant, then system (2.2) is completely decoupled and can be solved explicitly. Firm j grows exponentially at rate π_j. The industry is eventually dominated by those firms whose profitability equals $\max_j\{\pi_j\}$, in the sense that s_j approaches 0 if π_j is less than the maximum.

Just as in other biological contexts, however, growth is restrained by various negative feedbacks. In our model we will assume that instead of remaining constant, the price of the output may fall as supply increases and that the wages of the inputs rise as demand increases. Thus, we assume

$$\begin{aligned} &p \quad \text{is a function} \quad p(|x|) > 0 \text{ with } p' \leqq 0. \\ &w_i \quad \text{is a function} \quad w_i(a_{ix}) \geqq 0 \text{ with } w'_i > 0. \end{aligned} \tag{2.5}$$

In particular, the wage w_i is positive if the demand a_{ix} is positive. The condition $w_i(0) = 0$, while permissible, is only appropriate for factors used exclusively by the industry we are considering.

An individual firm's behavior is myopic if it disregards the impact of its own growth upon prices. To a potential monopolist this is the form of the problem of the commons which usually appears in a multi-player version of Prisoner's Dilemma. For the consumer the result is a healthy growth of output due to competition.

While the system can no longer be solved explicitly, we will describe its behavior in considerable detail, reserving the more intricate proofs for the following section.

First, we impose some technical assumptions:

For every j, there exists i such that $a_{ij} > 0$. As a_{ix} approaches infinity so does $w_i(a_{ix})$. For every j, $p(0) > \Sigma_i\, w_i(0)a_{ij}$. (2.6)

The first assumption states that any mode of production requires some input. The second statement is a mild description of the limitation of factor resources. Together they imply that total output remains bounded. The third assumption simply restricts our attention to firms whose technology is profitable at some total output level, which may be quite low. Our model is now (2.2) with assumptions (2.5) and (2.6).

1. Lemma: Assuming (2.5) and (2.6), there exist constants $K, \varepsilon > 0$ such that total production $|x| \geqq K$ implies that mean profitability $\bar{\pi}$ is negative. Hence, if $|x| \geqq K$, $d|x|/dt < 0$. On the other hand, $|x| \leqq \varepsilon$ implies $\bar{\pi}_j$ is positive for all j. Hence, for $|x| \leqq \varepsilon$, $dx_j/dt > 0$ if $x_j > 0$.

Proof: The state vector x can be written as $|x|\, s$ where s is the distribution vector of relative size and so

$$\bar{\pi} = p(|x|) - \Sigma_i\, w_i(|x|\, a_{is})a_{is}.$$

Now for any distribution vector s, some a_{is} is positive by (2.6) and so $\max_i(a_{is})$ is a positive continuous function as s varies over the compact set of distributions. So $\delta > 0$ exists such that $\max_i(a_{is}) > \delta$ for all s. According to the second condition in (2.6), we can choose $K \geqq 1$ such that $w_i(K\delta)\delta > p(1)$ for all i. Now, given x with $|x| \geqq K$ choose i_0 as a value of i at which the $\max_i a_{is}$ is attained for this particular distribution vector. The monotonicities of (2.5) imply

$$\Sigma_i\, w_i(|x|\, a_{is})a_{is} \geqq w_{i_0}(|x|\, a_{i_0s})a_{i_0s} \geqq w_{i_0}(K\delta)\delta > p(1) \geqq p(|x|).$$

So $\bar{\pi} < 0$ at x. Consequently, $d|x|/dt = |x|\,\bar{\pi} < 0$.

The third condition of (2.6) implies that there exists $\delta_1 > 0$ such that $p(\delta_1) > \Sigma_i\, w_i(\delta_1)\, a_{ij}$ for all j. Choose $\varepsilon \leqq \delta_1$ and positive such that $|x| \leqq \varepsilon$ implies $a_{ix} \leqq \delta_1$ for all i, e.g. $\varepsilon = \delta_1/\max(1,a_{ij})$. When $|x| \leqq \varepsilon$:

$$p(|x|) \geqq p(\delta_1) > \Sigma_i\, w_i(\delta_1)a_{ij} \geqq \Sigma_i\, w_i(a_{ix})a_{ij}.$$

So $\pi_j > 0$. Clearly, $x_j > 0$ then also implies $dx_j/dt = x_j\pi_j > 0$. #

(N.B. We use # to signal the end of a proof or the end of a statement whose proof is deferred)

It follows from lemma 1 that any solution path of the system eventually enters the compact region of states with total output at most K. Once in this region the path remains there. This means that the price feedbacks are sufficient to exclude the unrealistic possibility of unbounded growth.

The role of entry requires elaboration. If the current industrial state is x, then the support of the state, denoted supp(x), consists of those firms which are producing positive output. In general, for any vector $Y \in R^n$, supp $(Y) \equiv \{j \in J: Y_j \neq 0\}$. So if j is not in supp(x) $x_j = 0$ and firm j is not producing, i.e. its technical possibility is not currently being realized). If $\breve{J}$ is a subset of J we denote by $R^{\breve{J}}_+$ those states with support contained in $\breve{J}$, i.e. $R^{\breve{J}}_+ = R^{\breve{J}} \cap R^n_+$ where

$$R^{\breve{J}} = \{Y \in R^n : Y_j = 0 \text{ if } j \notin \breve{J}\} = \{Y \in R^n : \text{supp}(Y) \subset \breve{J}\}.$$

A state x is called interior if supp(x) = J; i.e. $x_j > 0$ for all j and all firms are in production. The set of interior vectors is denoted $\mathring{R}^n_+$, which is the interior of R^n_+ in the vector space R^n. The set of states with support equal to $\breve{J}$ is denoted $\mathring{R}^{\breve{J}}_+$, the interior of $R^{\breve{J}}_+$ in the subspace $R^{\breve{J}}$.

Along a solution path x_t of the system, the support does not change. If $x_j = 0$ at $t = 0$ then $dx_j/dt = 0$ and so x_{tj} remains at 0. If $x_j > 0$ at $t = 0$ then $x_{tj} > 0$ for all t; though it is possible that $\lim_{t \to \infty} x_{tj} = 0$. We regard the latter as the elimination of firm j by competition.

A state e is called an equilibrium if the solution path starting at e rests there constantly; i.e., $dx_j/dt = 0$ at e for all j. This means that for every j either $e_j = 0$ or

$\pi_j(e) = 0$. In other words, at equilibrium the profitability is zero for all firms which are currently in production. In particular, at an interior equilibrium $\pi_j = 0$ for all j.

Thus, while the asymptotic elimination of existing firms may occur in the model, the dynamics do not allow for entry by new firms. Instead we regard entry by firm j as a perturbation - a little exogenous jump to a new nearby state with x_j positive but small - after which the differential equations again apply to yield a solution path starting from this new initial point.

These considerations split the concept of the stability of an equilibrium into two parts. An equilibrium is internally stable if it is stable against perturbations preserving support. It is stable if it can withstand all perturbations. By stability against a perturbation from the equilibrium, e, to a nearby point, x, we mean that the solution path beginning at x remains near e. So if the outputs of the existing firms are altered slightly from their values at an internally stable equilibrium, e, the system may respond by continuing to change but the resulting outputs all remain close to their original values. We shall demonstrate that all equilibria are internally stable. However, they will not all be stable because some will be vulnerable to entry. Suppose $e_j = 0$ while x_j is positive though small; (i.e. firm j has begun production). Stability requires that x_j returns toward zero or at least remains small. However, if firm j is potentially superior (i.e. $\pi_j > 0$ can occur because $e_j = 0$) the industry may start at x and then move away from e toward a position where firm j produces a substantial share of the output.

Our analysis proceeds from the existence of a suitable Lyapunov function. Define:

$$P(|x|) = \int_0^{|x|} p(u)du, \quad W_i(y) = \int_0^y w_i(u)du$$

$$U(x_1,\ldots,x_n) \equiv P(|x|) - \Sigma_i W_i(a_{ix}). \tag{2.7}$$

It is easy to check that

$$\partial U/\partial x_j = p(|x|) - \Sigma_i w_i(a_{ix})a_{ij} = \pi_j \tag{2.8}$$

and so

$$dU/dt = \Sigma_j[(\partial U/\partial x_j)(dx_j/dt)] = \Sigma_j x_j \pi^2{}_j \geqq 0. \qquad (2.9)$$

Furthermore, $dU/dt = 0$ if, and only if, for all j, either $x_j = 0$ or $\pi_j = 0$; (i.e. if, and only if, the system is in equilibrium). Notice that if p and w_i are constant then $U(x) = p|x| - \Sigma_i w_i a_{ix} = |x| \bar{\pi}$, which is the total industrial profit.

U has an interesting economic interpretation. Fix the distribution vector s and let the total output $|x|$ vary so that we are moving along the ray through s towards x. Recall that $a_{ix} = |x| a_{is}$. Change the variables in the integral defining W_i by letting $u = va_{is}$, so that when $u = a_{ix}$, $v = |x|$. It then follows that

$$U(x) = \int_0^{|x|} [p(v) - \Sigma_i w_i(va_{is})a_{is}]\, dv. \qquad (2.10)$$

The integrand is now the mean profitability $\bar{\pi}$ evaluated at vs. So competititve dynamics act to maximize

$$U(x) = |x| \cdot \{1/|x| \int_0^{|x|} \bar{\pi}(v \cdot s)dv\}.$$

The expression in braces is an integral average of $\bar{\pi}$ along the ray from 0 to $|x|$. Presumably, if the industry behaved as a cartel the dynamic would tend instead to maximize $|x| \cdot \bar{\pi}(|x| \cdot s) = |x| \cdot \bar{\pi}(x)$, the total profit.

Still restricting our analysis to the ray defined by the fixed distribution vector s, we can draw the classical graph of unit income and cost for the industry as total output varies (see Figure 1). The graph reveals that U, the shaded area for output level $|x|$, is the Marshallian notion of surplus, being the sum of consumer surplus, factor surplus and total profit (labelled CS, FS and TP, respectively, in the figure). This interpretation of U is due to Sidney Winter.

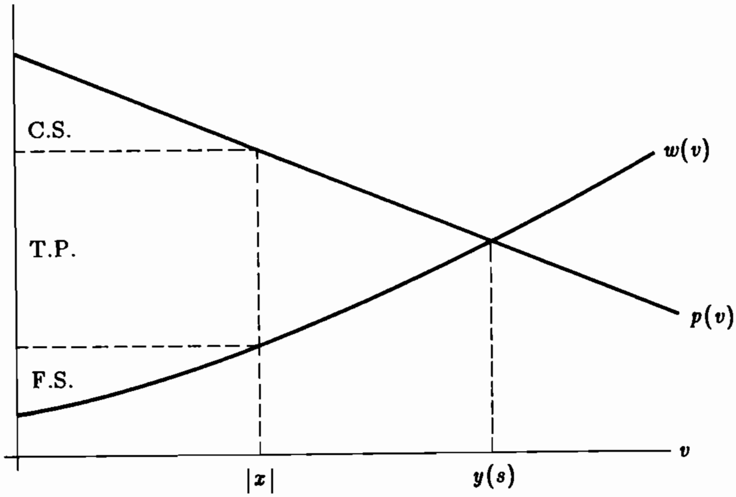


Figure 1

With $w(v) = \Sigma_i \, w_i(va_{is})a_{is}$,
$U(x) = U(|x|s) = C.S. + T.P. + F.S.$
C.S. = consumer surplus
F.S. = factor surplus
T.P. = total profit = $|x| \, \bar{\pi}(|x|s)$.

The output level where U achieves its maximum along the ray will be denoted y(s). Thus, at $x = y(s)s$ average profit $\bar{\pi}$ vanishes and thus, from (2.3), $d|x|/dt = 0$ at x . Lemma 1 and (2.10) imply that U increases when $|x| \leqq \varepsilon$ and decreases when $|x| \geqq K$. Thus, for every distribution vector s the zero profit level y(s) lies between ε and K. Observe that the condition $\bar{\pi}(x) = 0$ which defines $|x| = y(s)$ for $s = x/|x|$ is much weaker than the equilibrium condition $\pi_j(x) = 0$ for all j in the support of x.

Actually, U is not just a Lyapunov function. The dynamical system is really the gradient system for U when the notion of gradient is interpreted properly. So if we regard the graph of U as a topography lying over the multidimensional state space, the point representing the industry state climbs along the curve of steepest ascent, approaching equilibrium at a peak. The topography defined by U has a rather simple shape because U is a concave function.

Notice that the symmetric matrix

$$(\partial^2 U)/(\partial x_j \partial x_k) = p'(|x|) - \Sigma_i \, w'_i(a_{ix})a_{ij}a_{ik}$$

is negative semi-definite; i.e. if $Y \in R^n$ then

$$\Sigma_{j,k} \, Y_j \, [(\partial^2 U)/(\partial x_j \partial x_k)] Y_k = p'(x)(\Sigma_j Y_j)^2 - \Sigma_i \, w'_i(a_{ix})(\Sigma_j \, a_{ij} Y_j)^2 \leqq 0. \tag{2.11}$$

We denote by $H_x(Y)$ the expression in (2.11). The quadratic form H_x is called the <u>Hessian</u> at x of the function U.

If the initial point x has support equal to $\breve{J}$ then U is increasing on the solution path x_t, unless x is an equilibrium. Since the path remains in $\mathring{R}^{\breve{J}}_+$ the largest value it can approach is the maximum of the restriction of U to $R^{\breve{J}}_+$. We denote by $U^*_{\breve{J}}$ this maximum value and by $M_{\breve{J}}$ the set of points at which the maximum value is achieved:

$$U^*_{\breve{J}} = \text{maximum } U|R^{\breve{J}}_+$$

$$M_{\breve{J}} = \{x \in R^{\breve{J}}_+ : U(x) = U^*_{\breve{J}}\}. \tag{2.12}$$

In particular, when $\breve{J} = J$ we will drop the subscript and denote by U^* the maximum value of U and by M the set of points where it is achieved. Notice that if $\breve{J}$ is a proper subset of J there may be points x in $R^n_+ - R^{\breve{J}}_+$ at which $U(x) = U^*_{\breve{J}}$. By definition such points are not included in $M_{\breve{J}}$.

These maxima are all achieved because U(x) declines in value after x leaves the compact set defined by $|x| \leqq K$. Deferring the proof until the next section, we state our first major result:

2. <u>Theorem</u>: The set M of points at which U achieves its maximum consists entirely of stable equilibria and it includes all interior equilibria, if any exist. If x_t is any solution path in the interior $\mathring{R}^n_+$ then the limiting equilibrium $\lim_{t \to \infty} x_t$ exists and lies in M. #

3. <u>Corollary</u>: $\cup \{M_{\tilde{J}}: \tilde{J} \subset J\}$ is the set of all equilibria. For any solution path x_t the limiting equilibrium $\mathrm{Lim}_{t\to\infty} x_t$ exists. All equilibria are internally stable but only those in M are stable.

<u>Proof</u>: Restricting ourselves to those firms in some subset $\tilde{J}$ of J, we can apply the theorem to the system on $R^{\tilde{J}}_+$ and thus see that $M_{\tilde{J}}$ consists of equilibria including all of those in the interior $\mathring{R}^{\tilde{J}}_+$. Stability in $R^{\tilde{J}}_+$ implies internal stability with respect to the original system. Any solution path lies in some $\mathring{R}^{\tilde{J}}_+$ and so approaches an equilibrium in the corresponding set $M_{\tilde{J}}$.

But if e is an equilibrium not in M, i.e. $U(e) < U^*$, then we can perturb e slightly to a point x in $\mathring{R}^n_+$. The associated interior solution path x_t moves away from e because it approaches a point in M. #

The theorem and its corollary say that from any initial state the system goes to equilibrium. In essence, this is because the system admits the Lyapunov function, U. Furthermore, along a solution path in $\mathring{R}^{\tilde{J}}_+$, U rises toward its maximum value $U^*_{\tilde{J}}$. There are no saddle points or local maxima of intermediate height. This is a consequence of the concavity of U. Intermediate maxima only occur when we compare different supports sets $\tilde{J}$.

The approach to equilibrium would be easy to prove were the critical points of U – i.e. the equilibria – isolated. However, they need not be, and we want to see how to cope with these infinitely numerous degenerate equilibria.

In general, we call two states x and $\tilde{x}$ in R^n_+ <u>productively equivalent</u> if they determine the same total output and the same input demands, i.e. if $E^A(x) = E^A(\tilde{x})$ where $E^A: R^n \to R^{m+1}$ is the linear map defined by

$$E^A(x) = (-\Sigma_j x_j, \Sigma_j a_{1j} x_j, \ldots, \Sigma_j a_{mj} x_j). \tag{2.13}$$

The matrix of E^A is just a_{ij} augmented by an initial row of negative unit values. We denote by A the image of E^A, i.e., the subspace of R^{m+1} spanned by the columns of the's augmented matrix. The dimension of A is r ≡ the rank of the augmented matrix. We denote by B the kernel of E^A - i.e., the subspace of R^n perpendicular to the rows of the augmented matrix. Thus, states x and $\tilde{x}$, vectors of $R^n{}_+$, are productively equivalent if, and only if, the difference vector $x - \tilde{x}$ lies in B. Because the functions U(x), $|x|\bar{\pi}(x)$ and $\pi_j(x)$ depend on x only through the total output and input demands, they take on the same values at equivalent states. This proves the first part of:

4. Lemma (a) If x and $\tilde{x}$ are productively equivalent states then $U(x) = U(\tilde{x})$, $\bar{\pi}(x) = \bar{\pi}(\tilde{x})$ and, for all j, $\pi_j(x) = \pi_j(\tilde{x})$.

(b) If Y is a vector in B then, for all x, $H_x(Y) = 0$. Conversely, if for some x $H_x(Y) = 0$ then Y lies in B provided that $\Sigma_j Y_j = 0$. If either $p'(|x|) < 0$ or an interior equilibrium exists then the latter condition is superfluous.

Proof: That $H_x(Y) = 0$ for Y in B is clear from (2.11). Conversely, because $w'_i > 0$ and $p' \leqq 0$, $H_x(Y) = 0$ implies $p'(|x|)(\Sigma_j Y_j) = 0$ and for all i $\Sigma_j a_{ij}Y_j = 0$. Thus, $E^A(Y) = 0$ provided $\Sigma_j Y_j = 0$ which is true if $p'(|x|) < 0$. If an interior equilibrium e exists, then $\pi_j = 0$ for all j, i.e. $-1 = -\Sigma_i(w_i/p)a_{ij}$ for all j when w_i and p are defined by their values at e. This means that the first row of the augmented matrix is a linear combination of the remaining rows and so $\Sigma_j a_{ij}Y_j = 0$ implies $\Sigma_j Y_j = 0$. #

5. Corollary: (a) If x and $\tilde{x}$ are productively equivalent states with support contained in $\tilde{J}$, then $x \in M_{\tilde{J}}$ if, and only if $\tilde{x} \in M_{\tilde{J}}$.

(b) If e is an equilibrium and x is productively equivalent to e then x is an equilibrium provided that $supp(x) \subset supp(e)$.

Proof: (a) follows from $U(x) = U(\tilde{x})$. For (b) observe that $\pi_j(x) = \pi_j(e) = 0$ for all j in supp(e). Because this set includes supp(x), x is also an equilibrium. #

Notice that without the support condition part (b) need not be true because $\pi_j(x) = \pi_j(e)$ need not be zero when $e_j = 0$. Consequently the definition of $M_{\tilde{J}}$ had to be restricted to points with support in $\tilde{J}$.

E^A maps R^n_+ to a cone in A which we will denote by O. The interior states, i.e. $\mathring{R}^n_+$, consisting of strictly positive vectors, are mapped onto the interior $\mathring{O}$ of O in the subspace A. Thus each point of O represents an equivalence class of productive states. Each equivalence class is a convex cell consisting of the intersection of R^n_+ with a translate of the subspace B. A point lies in $\mathring{O}$ if <u>some</u> member of the equivalence class is interior. Except in the trivial case where B consists of zero alone, not <u>all</u> members of the equivalence class are interior. We provide the reader with an example of such a situation.

6. <u>Exercise</u>: If e is an equilibrium then there exists $\tilde{e}$ an equilibrium equivalent to e whose support, contained in supp(e), consists of at most rank(a_{ij}) firms. A fortiori, the number of firms in supp(e) is at most the number of factors m. (Hint: Choose $\tilde{e}$ equivalent to e such that $J = \text{supp}(\tilde{e}) \subset \text{supp}(e)$ is minimal among such points. Then prove $B \cap R^{\tilde{J}} = 0$). #

Although equivalent points have the same U value, the topography over an interior maximum is not like a hemisphere. Instead, imagine a cylindrical arch with lines drawn at constant height parallel to the cylindrical axis. The line segment at the peak of the arch represents an equivalence class of equilibria. The interior points of the segment are interior equilibria and each end point represents a boundary equilibrium with some firm excluded. We now slice the arch of this picture in such a way that each slice is an invariant manifold for the dynamics and each slice intersects the top segment at a single point. To do this we define for $b \in R^n$ the log-linear function:

$$L^b(x) = \Sigma_j b_j \ln x_j. \tag{2.14}$$

By summing over the j's in the support of b we see that L^b is defined and infinitely differentiable on the open subset

$$D_{\tilde{J}} = \{x \in R^n_+ : \text{supp}(x) \supset \tilde{J}\} = \{x \in R^n_+ : x_j > 0 \text{ for } j \in \tilde{J}\} \tag{2.15}$$

when $\tilde{J} = \text{supp}(b)$. Observe that for any $\tilde{J}$, $\mathring{R}^n_+ \subset D_{\tilde{J}}$. In fact, $D_{\tilde{J}} = \cup\{\mathring{R}_+^{J_1} : J_1 \supset J\}$.

7. Proposition: For $b \in R^n$ and $\tilde{J} = \text{supp}(b)$ let $x \in D_{\tilde{J}}$ and x_t be the solution path beginning at x (which remains in $D_{\tilde{J}}$ for all t).

(a) If b satisfies $E^A(b) < 0$ (i.e. the $m + 1$ components of $E^A(b)$ are nonpositive and are not all zero) then $L^b(x_t)$ is a monotonic increasing function of t. If e is the equilibrium $\text{Lim}_{t \to \infty} x_t$ then $e_j = 0$ for some j with $b_j < 0$. In particular, no equilibrium of the system lies in $D_{\tilde{J}}$.

(b) If b satisfies $E^A(b) = 0$ (i.e. b lies in B) then $L^b(x_t)$ is constant in t. Thus, the function L^b is an invariant of motion when the system is restricted to $D_{\tilde{J}}$.

Proof: As the support of x_t does not change as t varies, the solution path remains in $D_{\tilde{J}}$ and we can differentiate along the path:

$$dL^b/dt = \Sigma_j b_j \pi_j = p(|x|)(\Sigma_j b_j) - \Sigma_i w_i(a_{ix})(\Sigma_j a_{ij} b_j).$$

Furthermore, $p(|x|) > 0$ and for all i $w_i(a_{ix}) \geqq 0$ with equality only if $a_{ij} = 0$ for all $j \in \text{supp}(x) \supset \text{supp}(b)$. In the latter case $\Sigma_j a_{ij} b_j = 0$. In part (a), $\Sigma_j b_j \geqq 0$ and $\Sigma_j a_{ij} b_j \leqq 0$ for all i with at least one inequality strict. In part (b), all these expressions are zero. Hence the derivative along the solution path is positive in (a) and zero in (b). Thus, in (a) no equilibrium can exist in $D_{\tilde{J}}$. In particular, if e is the limiting equilibrium in (a), $e_j = 0$ for some j in supp(b). But if $e_j > 0$ whenever $b_j < 0$, then $L^b(x_t)$ would approach $-\infty$ as x_t approaches e. This is impossible because $L^b(x_t)$ is finite and increasing in t. #

Part (a) is useful in illustrating why interior equilibria need not occur. For example, we say that firm j_1 __dominates__ firm j_2 if $a_{ij_1} \leqq a_{ij_2}$ for all i with at least one inequality strict. This means that firm j_1 makes better use of all factors than j_2 and we would expect that from any initial state in which both firms occur, j_2 will be eliminated due to competition from j_1. Letting $b_{j_1} = +1$, $b_{j_2} = -1$ and $b_j = 0$ otherwise, $E^A(b) > 0$ and by part (a) $\text{Lim}_{t\to\infty} x_{tj_2} = 0$ for any solution path in $D_{\{j_1, j_2\}}$.

However, it is part (b) which provides the key result. For $b \in B$ we call $L^b(x)$ an __association parameter__ or association number of the interior state x. In general, if $b \in B^{\hat{J}} \equiv B \cap R^{\hat{J}}$ then the association parameter $L^b(x)$ is defined for any state x in $D_{\hat{J}}$, i.e. with $\text{supp}(x) \supset \hat{J} \supset \text{supp}(b)$. Part (b) says that the association parameters are constants of motion.

To interpret the association parameters, define for q any distribution vector (i.e. $q \in R^n_+$ and $|q| = 1$) the corresponding geometric average of a state x:

$$x^q \equiv \Pi_j x_j^{q_j}$$

where the product is taken over j in supp(q) or alternately over all j with the convention that $0^0 = 1$. Because $\Sigma_j b_j = 0$ for $b \in B$, we can multiply any $b \neq 0$ in B by a positive constant to obtain the sum of the positive coordinates unity. Then $b = q^+ - q^-$ where q^+ and q^- are distribution vectors of disjoint support and

$$L^b(x) = \ell n(x^{q^+}/x^{q^-}) \qquad (\text{supp}(x) \supset \text{supp}(b)) \tag{2.16}$$

where $\text{supp}(b) = \text{supp}(q^+) \cup \text{supp}(q^-)$. Thus $L^b(x)$ is the log of the ratio of geometric average output of two disjoint families of firms.

Recall that the dimension of A is r, the rank of the augmented production matrix. So n - r is the dimension of B. To obtain a representative list of association parameters, we choose a basis $\{b^1,...,b^{n-r}\}$ for B and define $L^B: \mathring{R}^n_+ \to R^{n-r}$ by:

$$L^B(x) = (L^{b^1}(x),..., L^{b^{n-r}}(x)). \tag{2.17}$$

8. Proposition: The function $E^A \times L^B$: $\mathring{R}^n{}_+ \to \mathring{O} \times R^{n-r}$ is a diffeomorphism, i.e. a one-to-one, onto differentiable function (recall that $\mathring{O}$ is the open cone $E^A(\mathring{R}^n{}_+)$). #

This result is proved in Akin (1979, page 81), and says that each interior state is described uniquely by its equivalence class and its association numbers. If we fix the association parameters, the interior state remains free to vary over a curved surface; technically a manifold of dimension r in $\mathring{R}^n{}_+$. The manifold intersects each interior equivalence class at exactly one point. The manifold is an invariant manifold for the dynamical system by Prop. 7(b) – i.e., a solution path beginning in the manifold remains in it.

We now state the result which completes the description of theorem 2.

9. Theorem: The set M of points at which U achieves its maximum is a single equivalence class with respect to productive equivalence.

Define J* to be the smallest subset of J such that $M \subset R^{J^*}{}_+$. Then $M \cap \mathring{R}^{J^*}{}_+ = M \cap D_{J*}$ is nonempty. In particular, interior equilibria exist (i.e. $M \cap \mathring{R}^n{}_+ \neq \emptyset$) if, and only if $J^* = J$, in which case $D_{J*} = \mathring{R}^n{}_+$.

With $B^{J^*} \equiv B \cap R^{J^*}$, the association parameters defined using vectors in B^{J^*} divide D_{J*} into invariant manifolds. If d is such a submanifold of D_{J*} then $d \cap M$ consists of a single equilibrium point e. e is globally asymptotically stable for the restriction of the system to d. Thus, if $x \in d$ and x_t is the solution path beginning at x, then x_t remains in d for all t and $e = \mathrm{Lim}_{t \to \infty} x_t$. #

While we again defer the proof, let us consider the special case where interior equilibria exist; i.e., U achieves its maximum at some point of $\mathring{R}^n{}_+$. By corollary 5, M is a union of equivalence classes: for $e \in M$, x equivalent to e implies $x \in M$. The new result is the converse which says E^A is constant on M. The point $E^A(M)$ lies in $\mathring{O}$ because interior equilibria exist. By proposition 8 we can regard E^A and L^B as providing a global nonlinear system of coordinates on $\mathring{R}^n_+$. So when we fix the L^B coordinates at arbitrary values we get a submanifold, d, of $\mathring{R}^n_+$ of dimension r. Since M is obtained by a particular choice of the E^A coordinates it is clear that $M \cap d$ consists of a single point e. By proposition 7 the L^B coordinates are invariants of motion and so d is an invariant manifold. On d, U is still a Lyapunov function but now e is the unique point at which U assumes its maximum value. Thus, by restricting to d, the degenerate nature of the equilibrium is eliminated and by Lyapunov's Theorem e is locally stable in d. Furthermore, theorem 2 says that the limit equilibrium of every solution path x_t in d lies in M. So $\mathrm{Lim}_{t \to \infty} x_t = e$. These results clarify the nature of the stability of e. If we perturb e to a nearby point x then x lies either on d or on a nearby d'. The solution path x_t moves back toward M in d' approaching $e' = d' \cap M$, an equilibrium near e.

It has probably occurred to the reader that it should be possible to avoid these equivalence class problems by aggregation. For example, if $a_{i1} = a_{i2}$ for all i then $b = (1,-1,0,...,0)$ lies in B and $L^b(x)$ is the log of the ratio x_1/x_2. We can define $y = x_1 + x_2$, define the new industrial state by $(y, x_3,...,x_n)$, and we get $dy/dt = y\,\pi_1$. However, this case of replication of identical firms appears to be the only situation in which aggregation is possible. In general, the problem is that while $d|x|/dt = |x|\bar{\pi}$ depends only on the equivalence class $E^A(x)$, the change in the demands a_{ix} depends not only upon $E^A(x)$, but on the association parameters, too.

Returning now to the general case, suppose we begin at some equilibrium e which is vulnerable to entry. After entry has occured the industry evolves to a new inequivalent equilibrium e* where $U(e^*) > U(e)$. In comparing the new equilibrium with the old one, methods of comparative statics do not apply because we are not considering a single equilibrium in a system with parameters. Instead we are comparing two widely separated equilibria in a single system. Recall that equivalent equilibria have the same total output. It seems reasonable to conjecture that one result of a new burst of technology - if it is commercially successful - is an increase in output. Given a homogeneous wage structure, this conjecture is true.

10. Theorem: Assume that the wage functions are homogeneous and of some common

positive degree, i.e. $w_i(a_{ix}) = \alpha_i(a_{ix})^k$ with k and the α_i's being positive constants. If e and e* are equilibria such that U(e*) > U(e), then the total outputs satisfy |e*| > |e|.

<u>Proof</u>: For any fixed distribution vector s, we defined the zero-profit level of output y(s) by $\bar{\pi}(y(s)s) = 0$. As we move out along the ray determined by s, the point $x = y(s)\cdot s$ is that level of output where $\bar{\pi} = 0$, or equivalent, $(d|x|)/dt = 0$. So y(s) is the real number y which satisfies the equation:

$$p(y) = \Sigma\ w_i(ya_{is})a_{is}.$$

Because the function p is nonincreasing and positive while each w_i is increasing and satisfies (2.6), the solution exists and is unique (see Figure 1). By our homogeneity assumption we can write this equation as

$$yp(y) = \Sigma_i\ \alpha_i(a_{ix})^{k+1} \tag{2.18}$$

with x = ys.

Clearly at any equilibrium $e = |e|\ s$, $\bar{\pi} = 0$ and so $|e| = y(s)$. So it suffices to show that $U(y(s^*)s^*) > U(y(s)s)$ implies $y(s^*) > y(s)$.

Integrating the homogeneous terms in the definition of U we get for all x:

$$U(x) = \int_0^{|x|} p(u)du - \Sigma_i\{(\alpha_i(a_{ix})^{k+1}) / (k+1)\}.$$

In particular, when x = ys with y = y(s), (2.18) implies:

$$U(ys) = \int_0^{y} p(u)du - \{(yp(y))/ (k + 1)\}.$$

Thus, in the homogeneous case, U(y(s)s) is a function of y = y(s) above. Furthermore, it is an increasing function of y because:

$$d/dy\ [\int_0^y p(u)du - (yp(y))/(k+1)] = (k/k+1)p(y) - (1/(k+1))yp'(y) > 0.$$

Thus, $U(y(s^*)s^*) > U(y(s)s)$ implies $y(s^*) > y(s)$. #

At first glance this result appears satisfying, but it turns out to be an artifact of the homogeneity assumption. To see this let us consider the two-firm-two-factor case. If neither firm dominates the other – necessary for the existence of an interior equilibrium – then we can label the factors so that (excluding the exceptional cases where $a_{12} = 0$ or $a_{21} = 0$):

$$a_{11} > a_{12} > 0 \text{ and } a_{22} > a_{21} > 0. \tag{2.19}$$

11. <u>Theorem</u>: Suppose that the production coefficients (a_{ij}) for a two-firm-two-factor case satisfy (2.19). Suppose further that the product price $p > 0$ is a constant. We can choose positive constants β_i, α_i ($i = 1,2$) to define the wage functions $w_i(d_i) = \beta_i(\alpha_i d_i + d^2_i)$, where d_i is the demand for factor i, so that the following conditions hold:

(1) Each single firm equilibrium is vulnerable to entry by the other.
(2) The unique interior equilibrium has total output less than the output at the firm 1 equilibrium.

The result is robust in the sense that slight perturbations of the price functions preserve conditions (1) and (2). #

Notice that if the industry actually consists of two firms then our model is inappropriate, because the myopic behavior we have postulated would not occur under conditions of oligopoly.

Suppose, however, that "firm 1" represents the aggregate of a large number of firms each using the column 1 technology. Together they produce the "firm 1" equilibrium output. If firms employing type 2 technology now enter at any small positive level, then the industry evolves to the stable equilibrium where both modes are in operation. However at this new, technologically richer equilibrium, the total output is less that it was under the regime of type 1 technology alone.

The result can be described geometrically by using Figure 1. If we use two different distribution vectors s_1 and s^*, then the price curves p(v) are the same for both , but the wage curves $w(v) = \Sigma_i w_i(v a_{is})a_{is}$ will usually be different. In going from the case of firm 1, $s_1 = (1,0)$, to the interior equilibrium distribution, s^*, the original wage function is replaced by one which intersects the p line earlier ($|e^*| < |e|$), but which encloses a larger total area ($U(e^*) > U(e)$). It is easy to draw such alternative wage curves, but that such alternatives can actually occur in the model requires proof.

Proof: First we change the coordinates so that total production is one of the new coordinates i.e. we define $u = x_1 + x_2$, $v = x_2$. As shown in Figure 2, the first quadrant in the x_1x_2 plane (i.e., R^2_+) is mapped on to the triangular region defined by $u \geqq v \geqq 0$. The x_1 axis ($x_2 = 0$) is now the u-axis and the x_2 axis ($x_1 = 0$) is the line u = v. The Lyapunov function U in the new coordinates is given by

$$U(u,v) = pu - W_1[a_{11}u-(a_{11}-a_{12})v] - W_2[a_{21}u+(a_{22}-a_{21})v] \quad \text{with } W'_i = w_i. \qquad (2.20)$$

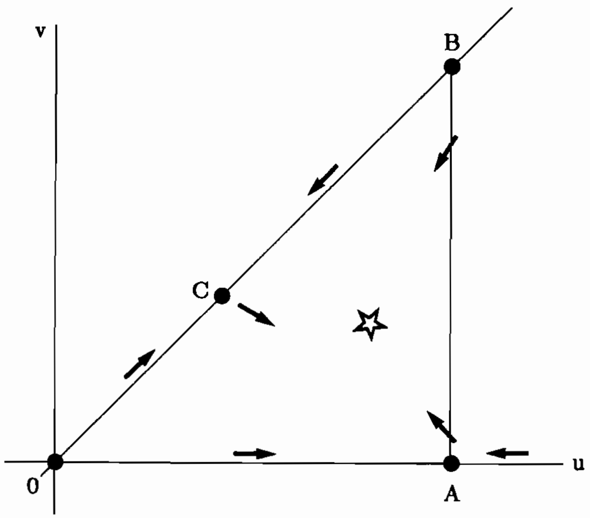


Figure 2

We focus on the right angle triangle OAB shown in Figure 2, where A = (1,0) and B= (1,1). The conditions which U must satisfy are:

(a) Along the segment OB, $\partial U/\partial v < 0$

(b) At the point A, $\partial U/dv > 0$

(c) At the point A, $\partial^2 U/\partial u \partial v < 0$

(d) At the point A, $\partial U/\partial u = 0$

(e) Along the segment AB, $\partial^3 U/\partial u \partial^2 v < 0$.

To see that these conditions yield the results we want, note that (e) and (c) imply $\partial^2 U/\partial u \partial v < 0$ along AB. Together with (d) this implies

(f) Along the segment AB, $\partial U/\partial u \leqq 0$ (vanishing only at A).

Condition (d) says that A is firm 1's equilibrium with total output 1, i.e. the maximum of U in the u-axis (= the x_1 axis) occurs at A. Moving vertically from A, (b) implies that U increases, so that the equilibrium at A is vulnerable to entry. To find firm 2's equilibrium we move along the ray u = v (= the x_2 axis) and compute

$$\partial U/\partial x_2 = \partial U/\partial u + \partial U/\partial v. \tag{2.21}$$

By (a) and (f), $\partial U/\partial x_2 < 0$ at B and so the maximum of U along this ray occurs earlier at some point we have denoted by C in Figure 2. At C, $\partial U/\partial x_2 = 0$ and so by (2.21) and (a), $\partial U/\partial x_1 = \partial U/\partial u > 0$ at C, namely firm 2's equilibrium is vulnerable to entry. Thus, the maximum value of U on the triangle does not lie on either the hypotenuse or the base. From (f) we know it does not lie on the vertical leg of the triangle either, and so the maximum of U on the triangle occurs at some interior point (denoted * in Figure 2). As * is an interior local maximum for U, it is the unique internal equilibrium point. As it is inside the triangle, the total output (its u coordinate) is less than 1. We do not know, but merely presume, that the output at * is greater than the output at C (as shown in Figure 2).

A little elementary calculus shows that (a)-(d) are equivalent respectively to (a')-(d') (defining $\delta_1 = a_{12}/a_{11}$ and $\delta_2 = a_{21}/a_{22}$ and noting that (2.19) says $1 > \delta_i > 0$):

(a') $(w_2(ta_{22}))/(w_1(ta_{12})) > (a_{11} - a_{12})/(a_{22} - a_{21}) = (a_{11}/a_{22})((1-\delta_1)/(1-\delta_2))$

for all $1 \geqq t > 0$.

(b') $(w_2(a_{21}))/(w_1(a_{11})) < (a_{11}/a_{22})((1-\delta_1)/(1-\delta_2))$

(c') $(w'_2(a_{21})a_{21}) / (w'_1(a_{11})a_{11}) > (a_{11}/a_{22})((1-\delta_1)/(1-\delta_2))$

(d') $p = w_1(a_{11})a_{11} + w_2(a_{21})a_{21}$.

Finally, for (e) it is sufficient that:

(e') $w''_i > 0$ for $i = 1,2$.

Notice that all positive choices for α_i, β_i imply (e'). To translate the remaining conditions in terms of the α_i, β_i choices, we define

$F(t) = (ta_{22}+\alpha_2)/(\delta_1(ta_{12}+\alpha_1))$,

$G(t) = (\delta_2(ta_{21}+\alpha_2))/(ta_{11}+\alpha_1)$

$K = (\beta_1 a^2_{11}(1-\delta_1)) / (\beta_2 a^2_{22}(1-\delta_2)$.

Some algebra shows that (a')-(d') are equivalent, respectively, to

(a'') $F(t) > K$ for all $1 \geqq t > 0$

(b'') $G(1) < K$

(c'') $G(2) > K$

(d'') $p = \beta_1[\alpha_1 a^2_{11} + a^3_{11}] + \beta_2[\alpha_2 a^2_{21} + a^3_{21}]$.

Now we are ready to make our choices. Notice that in order for both (b'') and (c'') to hold, G must be an increasing function of t which requires that $\alpha_1 a_{21} > \alpha_2 a_{11}$. It then follows from (2.19) that (with $\alpha_1, \alpha_2 > 0$) $\alpha_1 a_{22} > \alpha_2 a_{12}$ and so F(t) is also an increasing function of t. Furthermore, it is easy to check that, because $\delta_1\delta_2 < 1, \alpha_2 a_{11} > \delta_1\delta_2\alpha_1 a_{21}$ implies $F(0) > G(2)$. We choose α_1 and α_2 so that the ratio satisfies:

$$a_{21}/a_{11} > \alpha_2\alpha_1 > \delta_1\delta_2\, a_{21}/a_{11}.$$

Then F(t) > F(0) > G(2) > G(1) for all $1 \geqq t > 0$. Next choose the ratio β_1/β_2 so that G(2) > K > G(1). For example, if we choose $\alpha_1 = a_{11}$ and $\alpha_2 = \delta_2\, a_{21}$ we get K = G(3/2) if

$$\beta_1/\beta_2 = (\delta^2{}_2(3+2\delta_2)(1-\delta_2)a^3{}_{22}) \,/\, (5(1-\delta_1)a^3{}_{11}).$$

These choices for α_1, α_2 and the ratio β_1/β_2 imply (a")-(c"). Finally, multiplying the pair $(\beta_1, \beta_2$ by a suitable positive constant we get (d"). #

3. DISCUSSION

In introducing the Lyapunov function, U, we remarked that system (2.2) is the gradient system for U. The usual notion of the gradient system for U defines the absolute rate of change dx_j/dt by the partial derivative $\partial U/\partial x_j$. But in our model it is the relative rate of change $(d\,\ell n\, x_j)/dt = (dx_j/dt)\,/\,x_j$ which is the partial $\partial U/\partial x_j$. The concept of the gradient depends not only upon the function, U, but also upon the choice of an inner product used to measure the length of vectors and the angles between vectors. The usual, absolute rate version is associated with the ordinary Euclidean measurement of length and angles in R^n. The relative rate version uses the Shahshahani metric introduced in Shahshahani (1979) and elaborated in Akin (1979). In contrast with Euclidean geometry, the Shahshahani metric is a kind of Riemannian metric meaning that the inner product between vectors depends on the point at which the vectors are based. As a general rule of thumb, the Shahshahani geometry rather than the Euclidean one is the natural geometry for dynamic problems where relative rather than absolute rates are the focus of attention.

Let us now look back to the original assumptions of the model. The assumption that p and w_i are functions of |x| and a_{ix} respectively ensured exactly that the system is the gradient of the single function U. Then, the monotonicity assumptions of (2.5) imply the concavity of U. However, even with p and w_i as arbitrary functions of x, the dynamics of (2.2) still preserve association parameters. This is important because the plausible conditions of (2.5) are really too simple. That w_i depends on the factor demand a_{ix} signifies that the supply of factor i depends only on the wage w_i and not on the wage of other factors.

Thus (2.2), if equipped with more realistic supply functions, loses its gradient character and (possibly) the ubiquitous convergence to equilibrium. But the constants of motion L^b remain and so the associated foliation by invariant manifolds also endures.

APPENDIX

PROOF OF THEOREMS 2 AND 9.

We will begin by summarizing our previous results about the system and then execute the proof in a series of eight steps.

The function U defined by (2.7) and (2.10) is a Lyapunov function; i.e., U is constantly increasing on solution paths, remaining constant only when the entire system is at an equilibrium. U is a concave function with negative semi-definite Hessian given by (2.11).

The concept of productive equivalence is defined using the linear map E^A of (2.13). U is constant on each equivalent class. The vectorspace B is defined to be the kernel of E^A so that two vectors are equivalent if, and only if, their difference is a vector in B.

If $\hat{J} \subset J = \{1,...,n\}$ then $B^{\hat{J}} \equiv B \cap R^{\hat{J}}$ is the set of vectors in B with support contained in $\hat{J}$. A vector b of $B^{\hat{J}}$ defines by (2.14) the association parameter $L^b(x)$ of a state x in $D_{\hat{J}}$ (c.f. (2.15)). These association parameters remain constant on any solution path of the system which begins in, and so remains in, $D_{\hat{J}}$. So fixing the association parameters at particular values defines a manifold in $D_{\hat{J}}$ which remains invariant under the system. Restricting the focus to $\mathring{R}^{\hat{J}}_+ \subset D_{\hat{J}}$, proposition 8 says that each equivalence class in $\mathring{R}^{\hat{J}}_+$ intersects each invariant manifold at exactly one point. Thus, the association parameters of a state x in $\mathring{\tilde{R}}^J_+$ together with its equivalence class, described by evaluating E^A at x, uniquely describe x.

Finally, we mention a technical trick we have used before (e.g. corollary 3). If we are beginning at a point x and investigating the behavior of the solution path beginning at x, we can assume x is interior; i.e. supp(x) = J. If not, we can restrict the focus to the dynamical system defined only for the firms in supp(x). We will refer to this device as <u>interiorization</u>.

<u>Step 1</u>: Suppose e is an equilibrium with supp(e) = $\tilde{J}$. Then $U(e) = U^*_{\tilde{J}}$ (c.f. (2.12)). $U(e) = U^*$ if and only if $\pi_j \leqq 0$ at e for all j in J.

<u>Proof</u>: We begin with the latter statement. If some $\pi_j = \partial U/\partial x_j > 0$ at e, then by increasing the j coordinate of e slightly we increase U and so U(e) is not the maximum value of U. Now suppose $\pi_j \leqq 0$ at e for all j. Because the function U is concave, the graph of U remains below its tangent plane at e:

$$U(x) \leqq U(e) + \Sigma_j(x_j - e_j)\pi_j$$

where the $\pi_j = (\partial U/\partial x_j)$'s are evaluted at e. Because e is an equilibrium with support $\tilde{J}$, $\pi_j = 0$ for all j in $\tilde{J}$. For the remaining j's $\pi_j \leqq 0$ by hypothesis and $x_j - e_j = x_j \geqq 0$. So the sum is nonpositive and $U(x) \leqq U(e)$. Hence $U(e) = U^*$.

For the first result, we can assume – by interiorization – that $\tilde{J} = J$. At an interior equilibrium $\pi_j = 0$ for all j and so $U(e) = U^*$ according to what we have already proved.

<u>Step 2</u>: If e is an equilibrium with support $\tilde{J}$ then e is locally asymptotically stable in the invariant manifold of $\mathring{\tilde{R}}^J_+$ defined by the association parameters of e.

<u>Proof</u>: Again we can interiorize and assume $\tilde{J} = J$. No vector (except 0) of B is tangent to the invariant manifold through e. Technically, this follows from proposition 8

that the function $E^A \times L^B$ is a diffeomorphism. Hence, by lemma 3(b) the Hessian H_x is negative definite when restricted to vectors tangent to the invariant manifold. So, the Lyapunov function U restricted to the invariant manifold has a strict local maximum at e. Local asymptotic stability at e follows from Lyapunov's Theorem (see, e.g. Hirsch and Smale (1974) page 193).

<u>Step 3</u>: The set $M_{\bar{J}}$ is the intersection of $R^{\bar{J}}_+$ with a single equivalence class. $M_{\bar{J}}$ consists entirely of equilibria and includes all equilibria with support equal to $\bar{J}$.

<u>Proof</u>: By interiorization we can assume $\bar{J} = J$. By corollary 4(a) M is a union of equivalence classes. Because U is a concave function the set $M = \{x: U(x) = U^*\} = \{x: U(x) \geqq U^*\}$ is a convex set and so is its image under the linear map E^A. Now if $J_1 \subset J$ it is sufficient to show that $M \cap \mathring{R}_+^{J_1}$ - if nonempty - consists entirely of equilibria and has as its image a single point under E^A. Then every point of M is an equilibrium and the image $E^A(M)$ is a finite set (at most one point for each subset J_1). As M is connected and nonempty $E^A(M)$ is a single point; i.e., M consists of a single class.

By interiorization we can also assume $J_1 = J$. If $e \in M$ is interior then because e is a local maximum for U on $\mathring{R}^J_+$, $\pi_j = \partial U/\partial x_j = 0$ for all j at e and e is an equilibrium. Proposition 7 implies that the restriction of E^A to the invariant manifold through e is a diffeomorphism onto the open cone $\mathring{O}$ in A. Because $E^A(M \cap \mathring{R}^J_+)$ is a convex set in $\mathring{O}$, its preimage is a connected set of equilibria in the invariant manifold through e. But by Step 2 any equilibrium in the invariant manifold is locally asymptotically stable and so is isolated. Thus, the set of such equilibria is at once connected and discrete. Hence, it is the single point $\{e\}$. Because every equivalence class intersects the invariant manifold through e, the only interior points in M are those equivalent to e.

The last sentence of Step 3 is a restatement of Step 1.

<u>Step 4</u>: Every solution path x_t approaches a limiting equilibrium point as t approaches infinity.

Proof: For any system of differential equations whose solution all remain bounded (c.f. lemma 1) we can define the limit point set Ω of the solution path x_t as the set of limit points of convergence sub-sequences $[x_{t_n}]$. In general, Ω is a compact, connected set invariant under the system. We have to show that Ω consists of one equilibrium point. Often Ω may consist of more complicated elements than mere equilibria. Limit cycles offer a well-known example. In our case, the Lyapunov function ensures that limit cycles do not occur. Since $U(x_t)$ is a bounded increasing function of t it has a limiting value $\hat{U}$. By continuity of U $U(e) = \hat{U}$ at every point e of Ω. If we now compute the solution path beginning at e, then we remain in Ω by invariance. So U remains constant at $\hat{U}$ on this new solution path and this only happens at equilibrium. Thus, every point e of Ω is an equilibrium. We are not quite done. We have shown that Ω is a compact, connected set of equilibria in which U is constant. In most systems equilibria are isolated, e.g. there are only finitely many, and from this it is easy to show that Ω consists of a single point. In our system there are equivalence classes, convex cells, consisting of equilibria. So more information is needed to show that a single limiting equilibrium point exists.

Assume – by interiorization – that the original solution path x_t is interior - i.e., lies in $\mathring{R}^n_+$. Now let $e \in \Omega$ and let $\breve{J} = \mathrm{supp}(e)$. Any vector $b \in B^{\breve{J}}$ defines an association parameter $L^b(e)$. Because $b \in B$, L^b is defined and remains constant on the original path x_t. Hence, $L^b(e)$ is this constant value. Thus, the association parameters of e are determined by x_t. On the other hand, by Step 3 there is at most one equivalence class of equilibria with support equal to $\breve{J}$. Because the equivalence class and the association parameters are determined, so is e. We have shown that for every subset $\breve{J}$ of J, Ω contains at most one point with support equal to $\breve{J}$. Thus, Ω is finite. But as Ω is nonempty and connected, it does indeed contain exactly one point.

Step 5: If x_t is a solution path and $\mathrm{supp}(x_0) = \breve{J}$, then $\lim_{t\to\infty} U(x_t) = U^*_{\breve{J}}$, i.e. the limiting equilibrium for the path is one where U achieves its maximum on $R^{\breve{J}}_+$.

Proof: Interiorize to assume $\breve{J}=J$. By Step 4, x_t approaches some equilibrium. It is sufficient to show that if e is an equilibrium with $U(e) < U^*$ then no interior solution path approaches e. Let J_1 = supp(e). By Step 1, there exists j such that $\pi_j >$ 0 at e (of course $j \notin J_1$ since $\pi_j = 0$ for $j \in J_1$). So e has an open neighborhood G at every point of which $\pi_j > 0$. Assume x_t is an interior solution path with limit e and so the path eventually enters G and remains there; i.e., there exists T such that $x_t \in G$ for t > T. Because x_t is interior and lies in G, $dx_j/dt = x_j\pi_j > 0$ at x_t for t > T. So for t > T, x_{tj} is monotonously increasing and so cannot approach the limit $0 = e_j$. This contradiction completes the proof of Step 5.

For the remaining steps we will require an extension to infinite values of the log-linear function L^b.

<u>Lemma 2</u>: For $b \in R^n$ let $\tilde{J}_+ = \{j: b_j > 0\}$ and $\tilde{J}_- = \{j: b_j < 0\}$ so that $\tilde{J} = \tilde{J}_+ \cup \tilde{J}_-$ is the support of b. Let $\bar{D}_{\tilde{J}_+}$ denote $D_{\tilde{J}} \cup D_{\tilde{J}_-} \supset D_{\tilde{J}}$. On $\bar{D}_{\tilde{J}}$, L^b is a continuous function to the compactified real line $[-\infty, \infty]$ defined by:

$$L^b(x) = \begin{cases} +\infty & \text{for } x \in D_{\tilde{J}_+} - D_{\tilde{J}} \\ \Sigma_{\tilde{J}} b_j \ell n\, x_j & \text{for } x \in D_{\tilde{J}} \\ -\infty & \text{for } x \in D_{\tilde{J}_-} - D_{\tilde{J}} \end{cases}$$

<u>Proof</u>: On $D_{\tilde{J}_+}$ the sum $\Sigma_{\tilde{J}_+} b_j \ell n\, x_j$ is a real-valued continuous function. If $x \in D_{\tilde{J}_+} - D_{\tilde{J}}$ then $x_j = 0$ for some j in $\tilde{J}_-$. So as y in $D_{\tilde{J}}$ approaches x, $\Sigma_{\tilde{J}} b_j \ell n\, y_j$ approaches $+\infty$. A similar argument (or the same one applied to -b) yields continuity at points of $D_{\tilde{J}_-} - D_{\tilde{J}}$.

#

<u>Step 6</u>: Given $\tilde{J} \subset J$, let $\tilde{J}^*$ be the smallest subset of $\tilde{J}$ such that $M_{\tilde{J}} \subset R^{\tilde{J}}{}_+{}^*$. If x_t is a solution path in $\mathring{R}^{\tilde{J}}_+$ with $e = \mathrm{Lim}_{t\to\infty} x_t$ then supp(e) = $\tilde{J}^*$. In particular, $M_{\tilde{J}} \cap \mathring{R}^{\tilde{J}}_+$ is nonempty.

<u>Proof</u>: By interiorization assume $\tilde{J} = J$. We show first that {supp(x): $x \in M$} is closed under union. It is then clear that J* is the largest member of the collection.

By step 3 M is a single equivalence class and so it is convex. If $e_1, e_2 \in M$ then $e_3 = 1/2(e_1 + e_2)$ lies in M and $supp(e_3) = supp(e_1) \cup supp(e_2)$.

Now suppose $e \in M$ with $supp(e) = J_1$, a proper subset of J*. We show that e cannot be the limiting equilibrium of any interior path x_t. Choose $e^* \in M$ with $supp(e^*) = J^*$. Since e* is equivalent to e, $b = e^* - e$ is a vector of B. In the notation of the lemma, $e \in D_{\hat{J}-}$ and because supp(e) is a proper subset of supp(e*), $e \notin D_{\hat{J}}$. By the lemma L^b is defined and continuous at e with $L^b(e) = -\infty$. If x_t is an interior solution path then $L^b(x_t)$ is constant at some real value. Since x_t as well as e lies in $\bar{D}_{\hat{J}}$ $\mathrm{Lim}_{t\to\infty} x_t = e$ would violate continuity of L^b at e.

<u>Step 7</u>: Define the equilibrium function e: $R^n_+ \to R^n_+$ by $e(x) = \mathrm{Lim}_{t\to\infty} x_t$ where x_t is the solution path beginning at x. $e^{-1}(M) = R^n_+ - \{R^{\hat{J}}_+ : U^*_{\hat{J}} < U^*\}$ is an open subset of R^n_+ containing $\mathring{R}^n_+$. The restriction e: $e^{-1}(M) \to M$ is a continuous retraction onto M, i.e. $e(x) = x$ for $x \in e^{-1}(M)$ if and only if $x \in M$.

<u>Proof</u>: e is well-defined by Step 4. By Step 5 $e(x) \in M$ if, and only if $U^*_{\hat{J}} = U^*$ where $\hat{J} = supp(x)$. Now let $\{x_n\}$ be a sequence in $e^{-1}(M)$ converging to a point x_∞ of $e^{-1}(M)$. We want to show that the sequence $\{e_n \equiv e(x_n)\}$ in M converges to the point $e_\infty \equiv e(x_\infty)$ of M. Because M is compact it is enough to show that the limit of any convergent subsequence of $\{e_n\}$ is e_∞. So we suppose that $\{e_n\}$ converges to e* and prove $e^* = e_\infty$. Since there are only a finite number of subsets of J, we can restrict to a subsequence if necessary to assume that all points of the sequence $\{x_n\}$ have the same support (possibly different from the support of x_∞). Finally, by interiorization we can assume that this common support set is J. So we have reduced to the case where $\{x_n\}$ is a sequence of interior points converging to $x_\infty \in e^{-1}(M)$ and $\{e_n\}$ is a sequence in M converging to e*. Since the solution path beginning at x_n converges to e_n, we can choose y_n on the solution path with distance less than n^{-1} from e_n. Then $\{y_n\}$ is a sequence of interior points converging to e*. Because x_n and y_n lie on the same

solution path they have the same association parameters.

We must show $e^* = e_\infty$. Because e^* and e_∞ lie in the equivalence class M the difference vector $b = e^* - e_\infty$ lies in B. (This is where we use $x_\infty \in e^{-1}(M)$.) Let $J_1 \equiv \text{supp}(e^*)$ and $J_2 \equiv \text{supp}(e_\infty) \subset \text{supp}(x_\infty) \equiv \tilde{J}$. Then $\text{supp}(b) \subset J_1 \cup J_2$. Let $L_n = L^b(x_n) = L^b(y_n)$. Now $(y_n)_j \to e^*_j$. In the notation of lemma 2, $e^* \in D^{\tilde{}}_{J+}$ and $e_\infty \in D^{\tilde{}}_{J-}$. e_∞ is the limit of the solution path x_{∞_t} beginning at x_∞. So $\text{supp}(e_\infty) \subset \text{supp}(x_\infty)$ and $x_{\infty_t} \in D^{\tilde{}}_{J-}$ for all t. By the continuity result of lemma 2 and the constancy of association parameters

$$L^b(e_\infty) = \lim_{t\to\infty} L^b(x_{\infty_t}) = L^b(x_\infty).$$

On the other hand, $\{x_n\}$ and $\{y_n\}$ are interior sequences with limits x_∞ and e^* respectively. So by lemma 2 again

$$L^b(x_\infty) = \lim_{n\to\infty} L^b(x_n) = \lim_{n\to\infty} L^b(y_n) = L^b(e^*).$$

(Recall that y_n is on the x_n solution path.) Hence , $L^b(e^*) = L^b(e_\infty)$ and $e^* \in D^{\tilde{}}_{J+}$, $e_\infty \in D^{\tilde{}}_{J-}$ imply that this common value is finite. Thus, $e^*, e_\infty \in D^{\tilde{}}_{J}$. This implies that e^* and e_∞ have the same support.

$$0 = L^b(e^*) - L^b(e_\infty) = \Sigma(e^*_j - e_{\infty_j})(\ln e^*_j - \ln e_{\infty_j})$$

where the sum is taken over the common support set. But if $u,v > 0$ then $(u-v)(\ln u - \ln v) \geqq 0$ with equality only when $u = v$ (because the log is increasing). Hence, $e^*_j = e_{\infty_j}$ for all j in the support, i.e. $e^* = e_\infty$.

<u>Step 8</u>: Every equilibrium is internally stable. An equilibrium e is stable if, and only if $U(e) = U^*$, i.e. $e \in M$.

<u>Proof</u>: If $U(e) < U^*$ then e is vulnerable to entry; i.e. if x is an interior point close to

e then $e(x) \in M$. So the solution path approaches an equilibrium inequivalent to e, and a fortiori does not remain near e.

If $e \in M$ then e is stable in a very strong sense. Notice that if S is any subset of M then $e^{-1}(S)$ is an invariant set containing S. In particular, if G is a subset of M, open in M and containing e, then $e^{-1}(G)$ is open in $R^n{}_+$ and contains e. For $\varepsilon > 0$ let G_ε consist of the points of M of distance $\leqq \varepsilon$ from e, so that G_ε is compact and its M interior contains e.

Because U is a continuous Lyapunov function $\{U \geqq U^* - \varepsilon\} \equiv \{x: U(x) \geqq U^* - \varepsilon\}$ is a closed invariant set containing e in its interior. For $\varepsilon > 0$ small enough this set is bounded and so is compact. Also, for $\varepsilon > 0$ small enough $U^*{}_{\bar{J}} \geqq U^* - \varepsilon$ implies $U^*{}_{\bar{J}} = U^*$ and so this set is contained in $e^{-1}(M)$. Then $e^{-1}(G_\varepsilon) \cap \{U \geqq U^* - \varepsilon\}$ is a compact invariant set containing e in its interior. Because the intersection of these sets as ε approaches 0 is $\{e\}$, it follows that if N is any neighborhood of e, there exists $\varepsilon > 0$ such that $N \supset e^{-1}(G_\varepsilon) \cap \{U \geqq U^* - \varepsilon\}$. Because the latter set is invariant, stability follows.

ACKNOWLEDGEMENT

This paper received partial support from NIH grant No 1 RO1 GM 32153-01 GEN.

REFERENCES

Akin, E., The Geometry of Population Genetics, Lecture Notes on Biomathematics No. 31, Springer-Verlag, New York, 1979.

Hirsch, M. and Smale, S., Differential Equations, Dynamical Systems and Linear Algebra, Academic Press, New York, 1974.

Nelson, R. and Winter S., An Evolutionary Theory of Economic Change, Harvard University Press, Cambridge, MA., 1982.

Shahshahani, S., 'A New Mathematical Framework for the Study of Linkage and Selection', American Mathematical Society, Memoir No. 211, 1979.

Chapter 6

Technical Progess, Capital Accumulation, and Effective Demand: A Self-Organization Model

G. SILVERBERG

1. INTRODUCTION

Concepts such as the representative firm and consumer, "perfect" competition and foresight, and market equilibrium still dominate economic theory. This has led, on the one hand, to "micro" economic models of the Walrasian type, and to "macro"economic models of steady state, full employment growth on the other. At the same time, it is clear to even the most casual observer of the real economic world that diversity of expectations and business strategies, market power, sectoral, technological and foreign trade and many other forms of disequilibrium have always prevailed. Can these be regarded as mere blemishes on an otherwise correct method of analysis, or do they change "the ball game" the economic theorist is up against in a genuinely fundamental way?

The answers of economists such as Milton Friedman - that the underlying evolutionary process in capitalist economies ensures that the results of static optimization theory are generally enforced - appear a little too glib (see in particular the discussion in Nelson and Winter, 1982). It is by no means clear that once diversity of expectations and behavior, dynamic disequilibrium, and basic uncertainty about the future are introduced, they will simply go away, even in the proverbial long run. The purpose of this paper is to formulate such an evolutionary economic process explicitly, and to bring to bear some of the analytic methods of the theory of self-organization to clarify the relationship between industry structure, technical progress, and market evolution.

The theory of self-organization is concerned with the coupled dynamics of subsystems which interact to produce a "field" at the macro level, and in turn are themselves influenced by this very field. (For an overview of the principal methods and applications within the expanding field of self-organization theory see e.g. Haken (1983), Ebeling and Feistel (1982), Nicolis and Prigogine (1977), and Prigogine (1976)). Under suitable constraints - external input of energy and competition for scarce resources - such a dynamic feedback structure can be shown to result in coherent behavior of the disparate components, and to be characterized by a small number of aggregate parameters.

Since Adam Smith, the problem of the emergence of coherent behavior from the uncoordinated pursuit of self-interest has been one of the central economic questions. The answer he gave - the invisible hand of the market - and its modern generalization in general equilibrium theory, are among the first formulations of feedback structures in modern science. However, we now know that the conclusions to which this tradition leads - stability, uniqueness, and Pareto optimality of the resulting equilibrium - are only valid under severely restrictive and unrealistic assumptions.

In general, we must expect such systems to display much richer dynamics. The a priori assumption that an economy is actually governed by such an equilibirum can only be regarded as an article of ideological faith without scientific foundation. (Thus e.g. Smale (1980, p.289) objects that "general equilibrium theory has not successfully confronted the question 'How is equilibrium reached?' Dynamic considerations would seem necessary to resolve this problem.") Not only will the system display richer dynamics under given conditions, but it may also be subject to radical changes in regime due to innovations, switching effects, external shocks and internal noise. These changes, which may be likened to phase transitions, are the principal objects of study in the theory of self-organization. Moreover, they may well display a progressive or evolutionary character.

Since at least von Neumann and Morgenstern (1947, pp.8-12), it has been well known that the problem of economic decision making and rationality, even in the normative domain, is not simply that of individual maximization. Although much theoretical progress has been made in game theory since then, and some very suggestive parables have been uncovered (such as Prisoner's Dilemma and its extension into the evolutionary domain by Axelrod (1984)), game theory has failed to replace more classical optimization approaches in positive and normative economics. Indeed, Schwartz (1965, pp.83-88) has argued that the original game-theoretic program was too ambitious. H.A. Simon's concept of bounded rationality acquires attractiveness here - that economic agents adhere to simple rules of thumb which have shown themselves to be more or less satisfactory in most circumstances (e.g. Simon 1959). How such rules become established, and why they may retain their validity for a certain time, only to be replaced or overthrown later remains uncertain.

Schumpeter's conception of the economic process as one of creative destruction may hold the key to understanding how the market process "optimizes" in an evolutionary sense in historical time. He saw most economic agents as indeed acting in accordance with perceived rules of behavior. However, every now and then more enterprising individuals or firms attempt some form of innovation, whether technological, organizational, marketing, etc. Some of these innovations succeed, and as they acquire weight they force other agents to either bow out of the market, follow suit, or even "counter innovate" in their turn. In contrast to static neoclassical models, evolutionary processes of this kind require time - they are diffusionary in character. This implies that we are dealing with a very different form of optimization than in choice-theoretic models with their implicit infinite reaction speed: the process will be an ongoing one, open ended, of varying intensity, and possibly subject to dead-ends, breakdowns and other pathologies.

2. INDUSTRY STRUCTURE AND THE DYNAMICS OF COMPETITION

In a recent article, Kaldor (1983) points out that of the pioneers of the theory of effective demand, only H.-J. Rüstow took into account the fact that in a dynamic economy, the cost structure of industry will not be uniform, but rather will show a considerable variation both between firms and within the capital structure of firms themselves (see, for example, Rüstow 1951, 1984; the basic insight, however, goes back to Rüstow's 1926 doctoral dissertation.) This productivity gradient imposes a technological constraint on the realization of full employment, in addition to the distribution of income, the propensity to save, and the degree of monopoly power - the main factors otherwise associated with the problem of effective demand.

Numerous empirical studies have borne out that this gradient is a constant, and not merely a temporary feature of a dynamic economy. That this is so forces us to rethink our notion of market competition. For in a regime of "perfect" competition, a firm with a decided cost advantage should be able to corner the market instantly by slightly lowering its price below that of its competitors. On the other hand, the theory of imperfect or oligopolistic competition, while putting this notion into question, does not offer a clear alternative dynamic description.

If we turn to the businessman's view of reality as it comes across on the business pages of major newspapers, we immediately perceive that competition is regarded as a question of the struggle to defend or enlarge the firm's market share in the face of an overall demand largely beyond the firm's policy-making purview. This struggle, both domestically and internationally, is seen as turning on relative competitiveness, a concept which in practice is often left ill-defined, and until now has found no proper place in economic theory. That market shares display sufficient constancy to be regarded as key variables in strategic decision making indicates that many markets, particularly in manufacturing industry, are considerably less than "perfect". We suggest that market share is a key microeconomic variable of concern to the firm in determining output and pricing policy, instead of the marginalist concepts of the neoclassical theory of the firm for which there is little or no direct phenomenological evidence. This may be due to advertising, product differentiation, customer loyality, or simply the finite reaction time of the market to respond to price, quality and other signals.

Schwartz (1965, p.8) has shown that under such conditions of oligopolistic price competition , a Nash solution to the noncooperative game is mark-up pricing, with the mark-up factor dependent on the elasticity of demand facing each firm, itself a function of industry structure. Kalecki (1965, ch.1) also made mark-up pricing and monopoly power a central element in his trade cycle and distributional models, without however, being able to say what determines the mark-up factor.

As Schwartz's solution corresponds to a game-theoretical equilibrium, it does not show how this position is reached, nor how market shares may change over time as a result of the varying business strategies pursued by different firms. We propose the following first order model to capture the dynamics of "imperfect" market competition, based on the work of Manfred Eigen on prebiotic evolution (see Eigen

1971; for the analogy to processes of economic competition see Ebeling and Feistel, 1982, pp.247-8).

Let f_i denote the market share in total real orders of the i-th firm. Let E_i be the "competitiveness" of the i-th firm, $\bar{E}$ average industry competitiveness weighted by market share. For the time being we will not specify further what we mean by competitiveness. The evolution of market shares will be described by the following system of differential equations:

$$\dot{f}_i = A(E_i - \bar{E})f_i, \quad i = 1,n. \tag{1}$$

The coefficient A describes the speed of response of industry structure to disparities in competitiveness. "Perfect" competition would correspond to an infinite value of A. A might well depend on the degree of industry concentration, the extent of brand loyalty, or the like, and thus itself be subject to change as the industrial structure evolves (e.g. due to increasing concentration of the industry). In the following we will consider A to be a constant, however. As can readily be seen, the linear form of the equation for the rate of growth of market share ensures that the equations are consistent:

$$\Sigma \dot{f}_i = A\Sigma(E_i - \bar{E})f_i = A(\Sigma E_i f_i - \bar{E}) = 0 \tag{2}$$

as must be the case if the market shares are to continue to add up to one over time.

Equation (1) only describes the dynamics of market shares and leaves the overall level of demand D undetermined. Thus an additional equation is necessary to determine the absolute level of demand d_i facing each firm ($d_i = f_i D$). A typical candidate for modelling a small sector of the economy would be

$$D = D(\bar{p}, t), \quad \partial D / \partial \bar{p} \leqq 0, \tag{3}$$

where $\bar{p} = \prod_i p_i f_i$ is the average price. At a more aggregate level, demand might better

be represented as dependent on real wages in the consumer goods sector: $D = D(w/\bar{p})$.

If we regard the E_i's as constant, then the asymptotic behavior of the system is apparent. The firm with the largest E, say i*, will outgrow and displace all of the others until their market shares decline to zero and $\bar{E} = E_{i*}$. This process is independent of initial conditions, and is known as simple selection (see Ebeling and Feistel 1982, section 7.3, for an exhaustive mathematical discussion). Coexistence of more than one firm is only possible if surviving firms all have precisely the same competitiveness value. In a stochastic environment, this stringent condition is relaxed because small differences in the E's will not have a chance to lead to a systematic divergence of market shares. But of course this approach only becomes interesting when the E's are not constant, but subject to strategic planning and evolutionary selection. Thus it is necessary to model what constitutes competitiveness, and how it changes over time in response to the investment and pricing policies of the firm and their appropriateness in a changing market environment.

The first major constituent is obviously price, even in highly imperfect markets. The likelihood that a customer will switch his orders from one supplier to another is related not to absolute differences in price, however, but to percentual ones. Thus, price competitiveness can be represented to the first order by the logarithm of firm price, for example, with a negative sign to accord with the sign convention of equation (1) (lower than average price corresponding to higher than average competitiveness).

A second factor is quality competition, which for many products can outweigh the disadvantages of higher price. However, since it is not obvious which economic decisions determine the quality level, or whether it is subject to purely economic control at all, it is not clear how quality differences can be incorporated into such a model. In addition, after a certain point is reached, quality differentials transform themselves into the emergence of a new market and should perhaps better be regarded as a form of product innovation or differentiation. In view of these difficulties we will ignore the quality component in our definition of competitiveness.

Finally, the capacity constraint in the model of the firm's productive structure we introduce below must also be allowed to feed back on orders. Otherwise, a firm with a price advantage would continue to enlarge its market share even if it neglected to enlarge its capacity sufficiently to keep up with the expansion of demand. We model this aspect of competition by introducing a delivery delay, signaled to the market based on the backload of unfulfilled orders, the level of inventories, and the capacity constraint on production. Firms with an above-average delivery delay lose orders, even if they are price competitive.

Let dd_i be the current delivery delay of the i-th firm. The simplest way to combine its influence on overall competitiveness with that of price is to take a linear combination of the two:

$$E_i = -\ln p_i - a\, dd_i. \tag{4}$$

The reference, or average level of competitiveness, is given by

$$\bar{E} = \sum_i E_i f_i = -\Sigma f_i \ln p_i - a\, \Sigma f_i dd_i = -\ln (\prod_i p_i{}^{f_i}) - a\, \overline{dd}\,. \tag{5}$$

From equation (5) it is clear that the reference level for price competitiveness, if equation (1) is to be consistent, is not the arithmetic average $\bar{p} = \Sigma f_i p_i$ but rather the geometric average

$$\bar{p} = \prod_i p_i{}^{f_i}, \ (\Sigma f_i = 1). \tag{6}$$

3. EMBODIED TECHNICAL PROGRESS AND INVESTMENT STRATEGY: A VINTAGE APPROACH

It is widely recognized that investment in new equipment and plant is a major, if not the major, channel through which productivity improvements due to technical progress are realized. Other avenues certainly exist, such as "learning by doing"; the retrofitting of existing equipment with, for example, better instrumentation or other subsidiary devices, raising overall efficiency out of all proportion to the investment involved; better scheduling and control, etc. For the time being we will ignore these other factors and concentrate on pure embodied technical progress in which old equipment must be replaced by new equipment in order to raise productivity.

In a previous paper (Silverberg 1984) we introduced a model of the one-shot replacement of one capital structure by another in a dynamic setting. A choice-of-technique criterion was derived which is independent of momentary wage/price relations, and has the character of an evolutionary extremal function (for a discussion of evolutionary extremal principles see Eberling and Feistel, 1982 section 7.2). This criterion accords with Kaldor's stylized facts of economic growth, and makes Harrod neutrality of technical progress a plausible approximate result, instead of a strictly necessary presupposition of the model. Furthermore, the non-steady-state behavior of the model proved to be suggestive of a nonlinear superposition of long and short cycles.

As all practical managers know, however, technical progress in the modern era has been an ongoing, if perhaps variable affair, and not a one-time event. In the theoretical economic literature this fact was reflected in the vintage models (Salter 1962, Solow 1960, Kaldor and Mirrlees 1962). In the more management oriented literature, this was dealt with in the context of optimal replacement policy (see, e.g., Terborgh 1949, Smith 1961).

As has been pointed out by Clark (1980), however, a major drawback of the traditional vintage models has been their restriction to steady-state, full-employment equilibrium and a constant exponential rate of best practice technical progress. The management oriented analyses, such as the MAPI and payback period methods, are slightly superior in this regard: although they project a constant rate of technical progress into the future, they permit the current "vintage" under consideration to deviate from steady-state values (otherwise these methods would be of no practical usefulness). This difference in viewpoint also leads to an apparent contradiction between the two approaches. In the vintage models, a machine is replaced when its unit variable costs equal total costs of new machines (including depreciation charges and interest), i.e., when its quasi-rent falls to zero. In the optimal replacement literature, a machine is replaced when the discounted cost of not replacing it, including opportunity costs due to lost potential profits - given an expected rate of future technical progress - is minimized. It is easy to show that even in a golden-age steady state, they do not coincide, and the former is not in fact an optimal solution.

Moreover, the latter method makes at least some explicit provision for uncertainty about, and variation in, the rate of technical change. Thus expectational fluctuations could lead to real, and by no means irrational, investment fluctuations.

In an early article, Bain (1939) attempted to show how the resulting variability of the economic lifetimes of capital goods could lead to autonomous reinvestment cycles of a more convincing nature than that due to a simple "echo" effect. Rosenberg (1976) has also argued that technological expectations can lead to perverse patterns of optimum investment expenditures. Although the argument is only valid if technological change comes in discrete lumps (e.g. when IBM announces a superior new computer generation to be introduced in two years, and thereby undermines its own or its competitors' current sales), the point is well taken that the replacement decision is both an economic one and crucially based on expectations which need not, and often will not, be fulfilled. Moreover, since expectations about an uncertain future - expecially concerning the course of technological development - cannot in their nature be fully grounded rationally (Keynes' "animal spirits"), they will necessarily differ between entrepreneurs and be subject to imitational and state of confidence effects.

Thus in modelling the replacement decision, we adopt a rule of thumb perspective (the still widely employed payback period criterion) on the one hand and, in dealing with major innovations or "shifts in technological paradigms", a Schumpeterian pioneering entrepreneur perspective on the other. (This latter aspect is dealt with using an extension of the present model in Dosi, Orsenigo and Silverberg, 1986).

Let us begin with the normal business replacement decision. At any given time t, firm i disposes of capital stock $k_i(t,t')$, $T_i(t) \leqq t' \leqq t$, where k_i is the capital stock vintage introduced at time t', measured in capacity or efficiency units of the firm's product and $T_i(t)$ is the oldest vintage still in use at time t. Assume for the moment that the best practice technology is unambiguously defined at each moment in time, and is selected for investment by all firms. Thus the date of introduction of a vintage determines its operating characteristics (we ignore losses in productivity due to wear and tear, etc.), which we describe as follows. Prime labor required per efficiency unit is denoted by a(t), the ratio of overhead to prime labor required per efficiency unit by h(t), for equipment of vintage t. Let the nominal wage rate at time t be w(t), and the price per efficiency unit of best practice equipment be P(t). We could allow for other prime and overhead inputs such as materials, energy, etc., by making a, h and w vectors instead of scalars. We regard investment in fixed capital equipment as an irreversible decision and hence neglect any possible resale or scrapping value. We now posit that the decision to replace the equipment slice $k_i(t,t')$ by an equal capacity of new technology is governed by the following payback period calculation:

$$b_{tt'} = P(t)/\{h(t')a(t') + a(t') - h(t)a(t) - a(t)\} \cdot w(t) \leqq b_i \tag{7}$$

where b_i is the payback period demanded by the i-th firm. For the moment we allow firms to choose their own period arbitrarily in view of their expectations of the rate of technical progress and other factors. Later we will attempt to determine an evolutionary "optimum" value by searching for the investment strategy leading to the highest sustainable relative growth rate.

Thus given a (for the moment) fixed value b_i for the maximum payback period demanded of replacement investment, the firm will attempt to scrap a sufficient quantity of old equipment to maintain $b_{tTi} \leqq b_i$. For simplicity, we assume that the overhead labor/efficiency unit shows no distinct trend over time, i.e., h(t) = const. Solving equation (7) for T determines the desired scrapping margin

$$T_{bi}(t) = a^{-1}\{a(t) + P(t)/(b_i w(t)(1+h))\}. \tag{8}$$

If we denote by R(t) the amount of capacity which must be scrapped according to this criterion, then we have

$$R(t) = k_i(t, T_{bi}(t)) \cdot dT_{bi}/dt\ . \tag{9}$$

However, at any given time the firm may not be operating at its optimum scrapping margin, so that a dynamic behavioral rule must be specified to allow it to catch up (or fall behind) to this point. Thus we choose a first order control procedure:

$$dT_i/dt = \max(\mu(T_{bi} - T_i), 0). \tag{10}$$

We deal with the decision to expand capacity separately from the replacement decision. Theoretically, both should be dependent on a comparable expected rate of return calculation. In practice, this is easier said than done. The main obstacle is to estimate the gains and losses associated with different rates of growth of orders, as against capacity, including foregone orders due to a lack of capacity. The forecasting of orders is one of the variables most shrouded in uncertainty for the firm. It depends on the future state of the whole economy, the relative growth of the specific sector, and the competitive position of the individual firm. We deal with this problem by having recourse to a rule of thumb. Firms attempt to maintain their average rate of capacity utilization at a certain level u_o so as to have spare capacity available for demand peaks, and to have a buffer against forecasting mistakes. Given an initial forecast for the rate of growth of orders over the firm's investment planning horizon r, and an actual (smoothed) rate of utilization u, firms adjust their desired rate of expansion in response to the discrepancy between their actual and their desired rate of capacity utilization:

$$\dot{r}_d = \alpha(u - u_o). \tag{11}$$

Firms attempt to keep up with the expected growth of orders, correcting for an over or underutilization of capacity. This is plausible because with constant unit prime costs and fixed overheads, profits are an increasing linear function of the rate of utilization. On the other hand, the foregone profits due to limited capacity in a situation of high demand will also rise rapidly.

The rate of growth of orders confronting a firm is equal to the rate of growth of orders in the sector, plus the rate of growth of the firm's market share (growth rates are indicated by ˆ):

$$\hat{d}_i = \hat{D} + \hat{f}_i \tag{12}$$

For the time being we assume firms base their initial demand projections on "animal spirits", i.e. completely arbitrarily. The levels of expansion and replacement investment can now be determined as follows. Let K be total capacity. Desired net capacity change is $N = r_d K$ in capacity units and $r_d KP(t)$ in value units. This can be negative or positive. The desired level of replacement investment is R and RP in value units. If N is positive, then gross investment I = (N+R)P, and capacity S = R is scrapped and replaced. If N is negative - i.e., the firm deliberately shrinks to cut overhead expenses - then gross investment purely serves replacement needs.

However, the financial position of the firm may not permit this desired investment program to be realized exactly. The main source of financing for the modern corporation is its cash flow F. If the cash flow matches desired gross investment, I = F, then there is no problem. If I exceeds F then there is an incentive for the firm to seek external financing or tap its financial reserves (cash, draft accounts, securities). If F exceeds I the firm may prefer to go into financial investments, depending on the rate of interest. The question of external financing and investment introduces an additional complication which we postpone until section 6.

To simplify matters, we assume for the time being that firms always adjust their gross investment level to their cash flow, and ration funds proportionately between replacement and expansion investment as the need may be. This is consistent with the description of investment budgeting given by Cyert, Debrost, and Holt (1979) as well as interview studies such as Eisner (1956). We can thus define an investment realization factor $z = F/I$ if $F \geqq 0$ and $z = 0$ if $F \leqq 0$ to yield the following realized components of investment:

N	gross investment (value)	net expansion	replacement (in efficiency units)	scrapping
positive	z(N+R)P	zN	zR	zR
negative	zRP	N	zR	-N+zR

We can regard the evolution of the firm's capital stock as a process of age-dependent population dynamics. Gross investment I adds new equipment to the stock and scrapping S due to replacement and/or contraction decisions removes old equipment from the tail end (the age of which is endogenously determined and variable). Given a predetermined development pattern for best practice productivity, we can now ask how the productivity of the entire capital stock will evolve as a result of these investment decisions. For simplicity, we will regard overhead costs as unvarying ($h(t)$ = constant), and restrict the analysis to prime unit labor costs and productivity $a(t)$ and $g(t)$. Consider an arbitrary characteristic of technology $x(t)$ solely dependent on a vintage's date of introduction. Then we can define the average value of x over a given capital stock as

$$\bar{x} = 1/K \int_T^t x(t')k(t,t')dt'. \tag{13}$$

Let $B = I/P$ be gross investment in efficiency units. Then the net change in capacity is obviously

$$N = \dot{K} = B - S. \tag{14}$$

The rate of change of $\bar{x}$ can now be derived from equation (13) by differentiation:

$$d/dt\,(\bar{x}) = 1/K\ \{B(x(t) - \bar{x}) + S(\bar{x} - x(T(t)))\} \tag{15}$$

If we consider a growing firm so that $N = B - S \geqq 0$, then from our investment table above $B = N + R$, $S = R$, and thus

$$d/dt\,(\bar{x}) = 1/K\ \{N(x(t) - \bar{x}) + R(x(t) - x(T(t)))\}. \tag{16}$$

If we now return to prime labor productivity g and assume it is a monotonically growing function then $g(t) \geqq \bar{g}(t) \geqq g(T)$. Substituting g for x in the last equation, this implies that a dollar's investment will contribute most to raising average productivity if it is used to rationalize rather than expand the capital stock. Thus, if for some reason businessmen expect demand to stagnate or decline and are under competitive pressure, then they will have good reasons to prefer accelerated modernization - i.e., cost cutting - to net capacity expansion.

Equation (15) shows a certain affinity with Kaldor's technical progress function (Kaldor and Mirrlees 1962). Whereas Kaldor posited a relationship between the rate of growth of productivity and investment per man on new equipment, equation (15) relates the rate of growth of productivity of operating equipment to the level and composition of investment as well as the rate of development of best practice technology. The duality of the investment decision (replacement vs. expansion), and not just the level of gross investment, is thus shown to be a determining factor in the evolution of productivity.

Let us return to the expansion investment decision, the effect of which is to enlarge or shrink the productive capacity of the firm's capital stock. In section 2 we introduced the delivery delay as a component of market competitiveness. Here we have to specify the delivery delay dynamically as a result of net investment and the level of production decisions. Letting incoming orders be denoted by d and outgoing shipments by y we have

$$\dot{L} = d - y \tag{17}$$

where L is the backlog on the order book of the firm. The current delivery delay is then

$$dd = L/y. \tag{18}$$

We model the firm's short-term production decision by assuming that it attempts to maintain this delivery delay at some standard operating level dd_o. If the firm has spare capacity at its disposal, this is no problem. The level of production adjusts to demand with a certain finite reaction speed due to adjustment costs, the rate at which hiring and firing can take place, etc. The firm runs up against an operating barrier at full utilization of capacity, however. We model this process with the following equation:

$$\dot{u}/u = \beta(dd - dd_o)(1.1 - u^2), \text{ if } u \leqq 1 \;\Big|\; = 0 \text{ if } u \geqq 1, \; \dot{u} \geqq 0. \tag{19}$$

It would be more realistic to assume that shipments are from stock and that the level of production is also determined by a desire to maintain stock in a given relation to average incoming orders. This would necessitate an additional intervening stock adjustment equation, without fundamentally changing the mathematical structure of the system, so that we will ignore inventory dynamics for the present.

4. PRICE DYNAMICS

Equations (1) and (4) describe how quantities react to price and delivery delay signals. While the delivery delay can be determined by production level and capacity expansion decisions, it still remains to determine how prices are set. In the literature, there are several traditional approaches. The marginalist theory of the firm regards producers as having minimal market shares, and thus being absolute "price takers"; they produce up to the point at which marginal cost equals the uniform market price. Since marginal cost is constant and below average unit cost in our model (as most empirical studies have shown for manufacturing), this solution is clearly inapplicable. Nor is the monopoly price/output criterion - marginal revenue equal to marginal cost - applicable, because in oligopoly, firms no longer confront static individual demand curves.

The full cost, target price, or markup theory holds that firms determine average unit cost at some standard operating capacity, and add a fixed percentage to arrive at a price which is otherwise independent of demand and competitive pressures. Although markup pricing has seemingly been confirmed repeatedly in interview studies, there is a problem in applying it to our model.

We want to study markets with a differential cost structure resulting from different investment strategies. If firms apply a rigid (and uniform) markup to costs, their prices will also vary accordingly. High cost firms will therefore be condemning themselves to losses of market share. It may be in their longer term interests to avoid this by lowering the price. Low cost firms may be foregoing short term profits by charging too low a price. Furthermore, a firm may be tempted to expand its market share by price cutting, especially in periods of slack demand. Whether it will succeed, however, depends on whether its competitors can or are willing to follow suit. This obviously leads into the thicket of game theory, where neither a "kinked" demand curve nor a static payoff matrix apply, but rather a "dynamic" payoff function as given by equation (1).

A general solution is evidently out of reach. A simplification is possible, however, along lines sketched in a very suggestive paper by Williamson (1965). There he argues that oligopolistic markets are characterized by two relatively stable states: a state of high implicit inter-firm cooperation with high markups and little price competition, and a state of minimal cooperation and low profit margins resulting from price warfare. The dynamic system he proposes to capture this phenomenon can be seen in retrospect to be an example of a cusp catastrophe. We will restrict the following analysis to the high cooperation structure, and assume that prices change in a continuous and stable manner in response to slowing varying changes in competitive cost structures.

The problem of reconciling a markup approach with the tendency to a uniform price level still remains. One solution for which there is considerable empirical evidence is price leadership: one firm, usually the largest or oldest in the industry, sets prices according to a strict markup formula, and the others more or less completely adjust to this price. This would imply that the latter no longer operate according to a strict markup. Moreover, the question of singling out the price leader and explaining why this sometimes changes, still remains open. There are elements of a game and information theoretic nature here which would carry us too far beyond the confines of our present methodology. (Eiteman, 1960, has argued, however, that price leadership is assumed by the most "mechanized" firm, i.e., the one with the lowest unit variable and highest unit fixed costs.)

Instead, we propose the following dynamic adjustment equation which appears to capture the main aspects of the problem, and at the same time allows for shifts in the price structure due to long-term changes in relative cost competitiveness. Let p_i be the current price of the i-th firm and p_{ci} its desired markup price. Let $\bar{p}$ be the current average market price (weighted by market share). These two price "pegs" are conceived as acting as "centers of attraction" for the current price in the following manner:

$$\dot{\rho}_i = a_i(\rho_{ci} - \rho_i) + a_2(E_i - \bar{E}), \qquad (20)$$

where the ρ's are the logarithms of the corresponding price variables. A firm's price is in equilibrium when it is the geometric mean of its markup and the average market price. If we neglect any variance in the delivery delays:

$$p_i^{eg} = p_{ci}^{\lambda}\ \bar{p}^{\ 1-\lambda}, \quad \lambda = a_1/(a_1 + a_2), \tag{21}$$

as can be seen by setting equation (20) equal to 0, inserting equation (4) and averaging over market shares. The dominant firm's price will always have the greatest influence on the average price, and thus its markup price will indeed serve a price leadership function. On the other hand, a small low cost firm can slowly begin to pull the general price level down without departing so far from it as to initiate a price war. If we regard market shares as changing significantly more slowly than the adjustment of prices to their "equilibrium" values, then we can set $\dot{f}_i = 0$ to obtain

$$\begin{aligned} d/dt\,(\bar{\rho}) &= d/dt(\Sigma\ f_i\rho_i) \cong \Sigma\ f_i\dot{\rho}_i \\ &= \Sigma\ f_i(a_1(\rho_{ci} - \bar{\rho}) + a_2(\bar{\rho} - \rho_i)) \\ &= a_1(\bar{\rho}_{ci} - \bar{\rho}). \end{aligned} \tag{22}$$

This implies that the average price level adjusts to average costs with a certain (presumably high) characteristic speed. In "price equilibrium"

$$\bar{\rho} = \bar{\rho}_c \tag{23}$$

so that strict markup pricing holds in the aggregate, even if it does not precisely hold for each individual firm.

If we cease regarding market shares as constant, we can derive a more exact relation for the dynamics of the aggregate price level by continuing to assume that price dynamics are significantly more rapid than market share adjustments (the so-called adiabatic approximation):

$$d/dt\,(\bar{\rho}) = A\ \lambda\ \sigma_{\rho ci} + a\ A\ \lambda\ \langle(\overline{dd} - dd_i)\,\rho_{ci}\rangle + \langle\hat{c}_i\rangle. \tag{24}$$

The first term represents the shift to lower cost firms due to simple price competition. The second represents the shift to higher cost firms in periods of high demand due to their shorter delivery delays (assuming delivery delay and cost are normally inversely correlated). The third term represents the rate of increase of average costs.

5. SUMMARY OF THE MODEL

The following Flowcharts (Figures 1 and 2) connect together the various subsystems of the model described above. At this stage in the simulations, all reaction coefficients (e.g. for prices, production level, etc.), tradeoff factors (e.g. between price and delivery delay in equation (4)), and the markup factor are assumed identical for all firms in the industry. Although the demand growth "animals spirits" projections can be left open as an additional degree of freedom, we concentrate for simplicity on differences in the payback period (or equivalently in the desired maximum age of equipment) in the following discussion of strategic interaction.

6. NUMERICAL ANALYSIS OF SYSTEM BEHAVIOR

The above system was implemented as a model of sectoral behavior on a computer using a Runge–Kutta–Merson subroutine for the solution of nonlinear differential equations (NAGDO2BBE). The program allows for numerical experimentation with up to five different investment strategies simultaneously, and different exogenous rates of growth of overall demand, nominal wages and best practice productivity. Capital stock vintages are represented discretely by taking the average value of gross investment over short intervals (one-half year in the following simulations). Parameter values were initially selected to yield plausible time constants in the various reaction equations. A systematic sensitivity analysis will be performed in the near future.

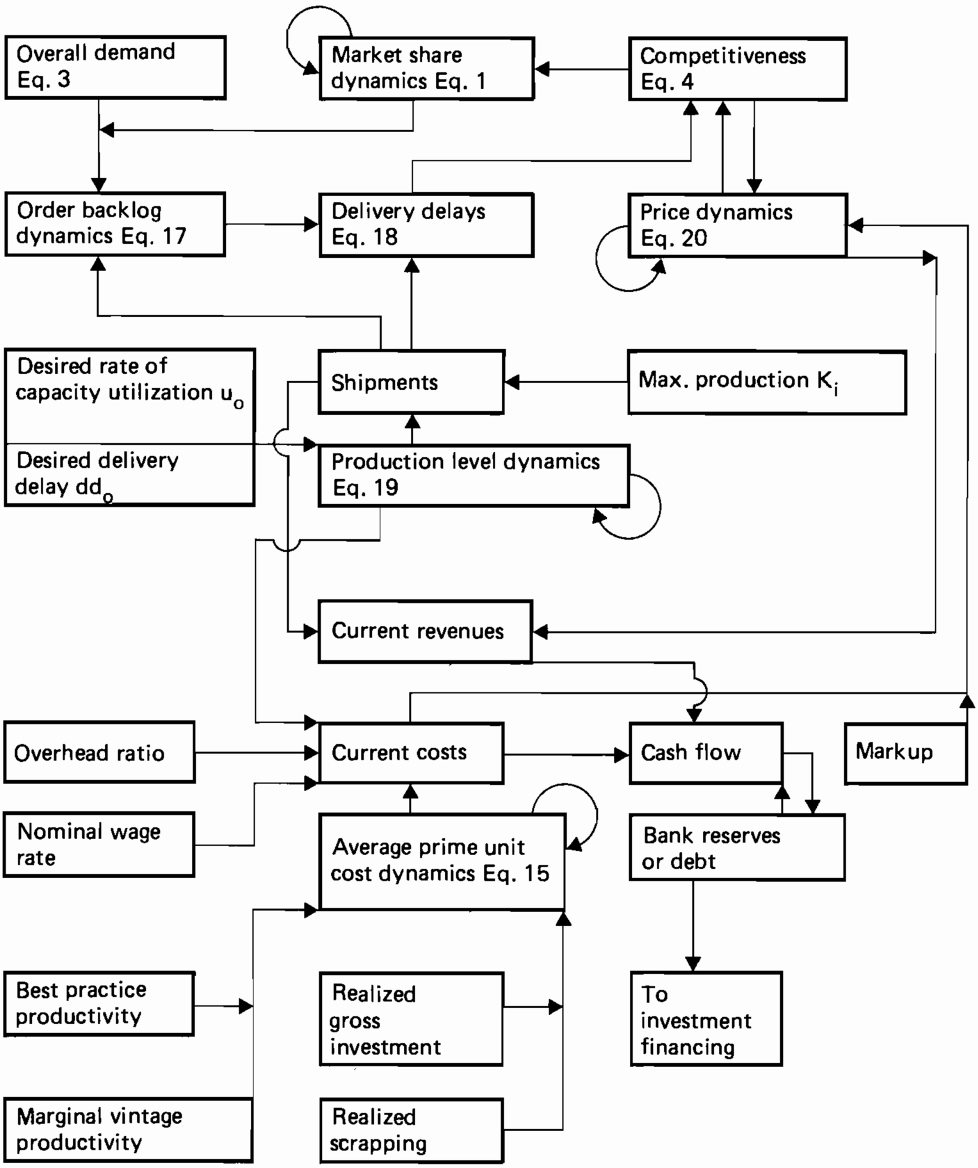


Figure 1 Dynamic Structure of Orders, Shipments, Prices, Costs, Delivery Delays, Productivity and Cash Flow.

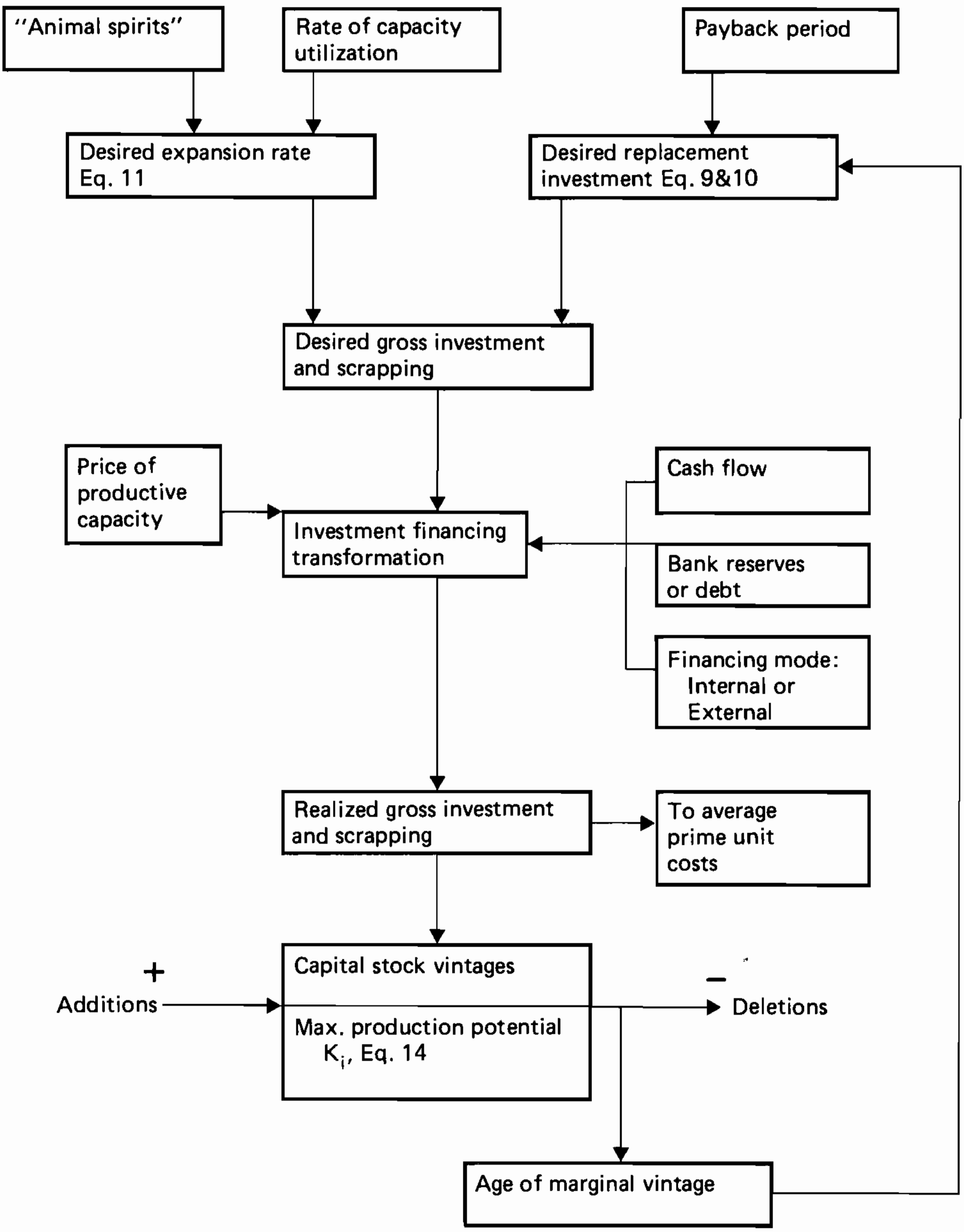


Figure 2 Dynamic Structure of Investment Plans, Financing Constraints, and Evolution of the Capital Stock.

The sectoral model can be run with two different financing modes in order to investigate the influence of financial constraints on evolutionary performance. In the first mode, firms can always realize their investment plans fully by borrowing on an unrestricted capital market. Since no limit is imposed on the extent to which firms can go into debt to finance investment, the "black hole" case of increasing debt ratios is not ruled out. This requires that a profitability criterion be imposed which may be at odds with the changes of relative market shares. Realistically, this should feed back in the model on the willingness of sources of finance to continue to lend to such firms. Since this would require a full-fledged model of capital markets, however, this question was not pursued further along these lines. Instead, another financial mode was introduced to characterize internal financing through cash flow and liquid reserves. In this case, if cash flow exceeds desired gross investment, investment plans are fully realized and the difference accrues to the liquid account (which earns interest). In the reverse event, where cash flow falls short of desired gross investment, investment plans can still be fully realized if sufficient funds are available from the liquid account. If not, investment is rationed, as described above, so that it can just be financed by current cash flow. This mode provides for an automatic performance feedback on investment while allowing the pattern of receipts and expenditures to differ over time.

In the following, two runs with essentially identical initial conditions, but differing in investment mode, will be described. In the first run with unlimited external financing, five firms or blocks of identical firms compete for market share with one another, differing only with respect to their replacement policies. Each firm replaces its oldest vintage when it has reached a specified age (here 10, 13, 16, 19, and 22 years). This is equivalent to differing payback periods when best practice productivity is increasing at a constant exponential rate. The growth rate of overall demand is exogenously given (3%) and is equal to the initial growth rates of each firm. However, the rates change over time in response to variations in the rate of capacity utilization they experience. The firms start the run with identical market shares, prices, and capital stocks (a rectangular distribution spanning 20 years). It is apparent that the firm with the shortest desired lifetime for its capital stock thereby necessarily achieves a cost advantage over its competitors which translates into a constantly growing market share (see Figure 3). Whether this is a profitable

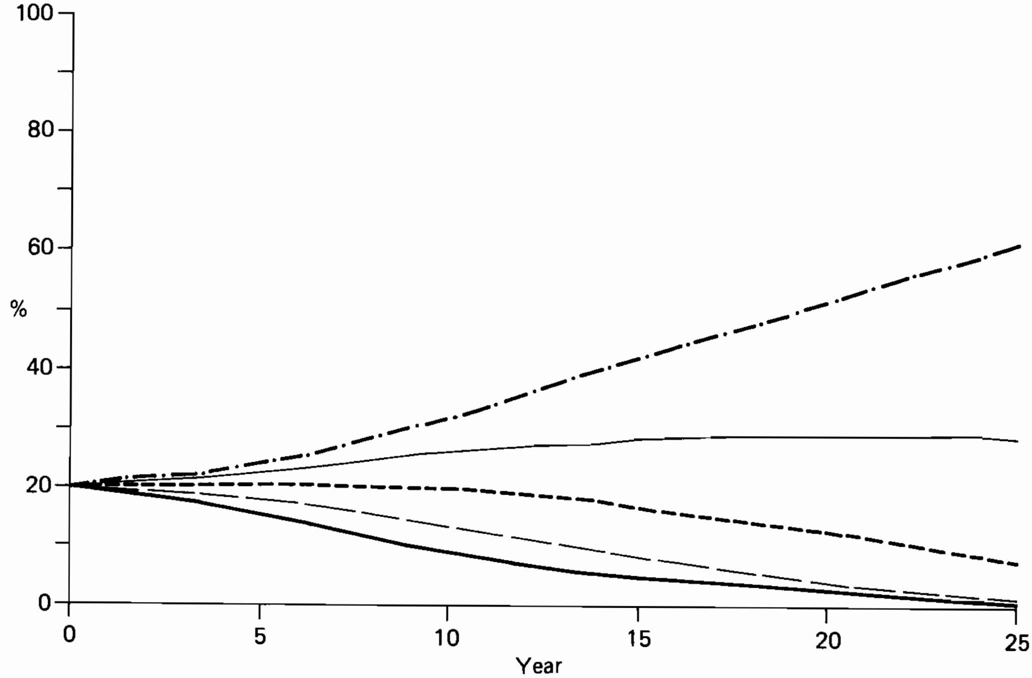


Figure 3 Evolution of Market Shares with Unlimited External Financing **(Run 2)**

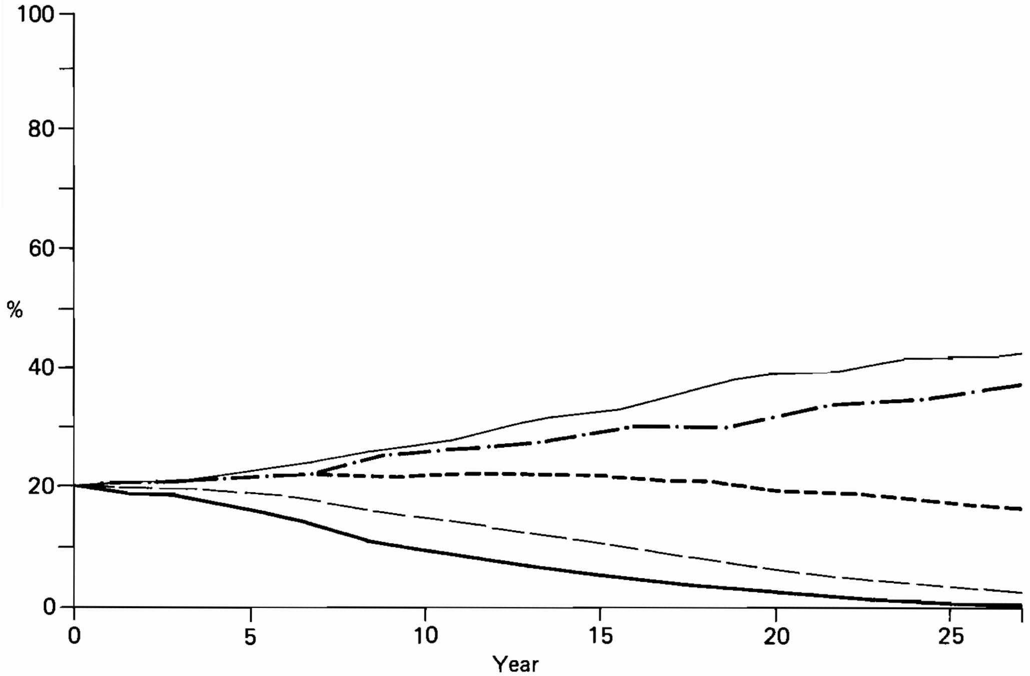


Figure 4 Evolution of Market Shares with Internal Financing **(Run 5)**

proposition is another question, however. Figure 5 shows the gross operating profit rate on physical capital (at replacement value) as well as the profit rate obtained by including interest income or expenditures, but net of depreciation (linear) in relation to all assets (i.e. including positive bank balances). It is clear that profitability may suffer from an overhasty replacement program.

Although these runs are purely hypothetical exercises in evolutionary "optimization" and are not meant, at this level, to describe a realistic market structure, they already illustrate a very significant point. The interaction of firms pursuing diverse investment strategies with differing vintage compositions of their capital stocks, results in clear non-steady-state behavior of firm-level and aggregate variables (in the present case with a periodicity of around 15 years), even though exogenously given variables were chosen precisely to permit such a growth steady state. This can be seen for the gross profit rate (Figure 5), delivery delays (Figure 8) and desired and realized capacity expansion rates (Figure 6).

The dynamics of price competition are shown in Figure 7, for the internal financing case. The price leadership function of the largest firms is apparent, as is the increasing inability of marginal firms to keep up with the market development due to rising costs (until a radical cutback in their capital stocks allows their prices to fall concurrently with their practical disappearance from the market).

The capital stock dynamics are among the most interesting features of the model. Figures 9 and 10 show the ratios of average prime labor per unit output to the best practice value, a surrogate for the average age of the capital stock for the two cases. It is apparent that these values fluctuate even though the exogenous parameters represent a growth steady state. To begin with, there is a sharp drop in this ratio as firms discard their older initial vintages to attain their maximum desired age. An "echo" effect in the form of a pronounced drop in the ratio can also be seen after 10, 13, 16, 19, and 22 years for firms 1 - 5 respectively. This is considerably less sharp the second time around for firm 1 (at year 20). This ratio is not only a function of the spread of the age distribution but also of its shape. This means that under non-steady-state conditions it will also depend on past growth rates as well: high growth rates in the past (other things being equal) leading to a younger capital stock. The significance of this fact can be seen by comparing Figures 9 and 10, which

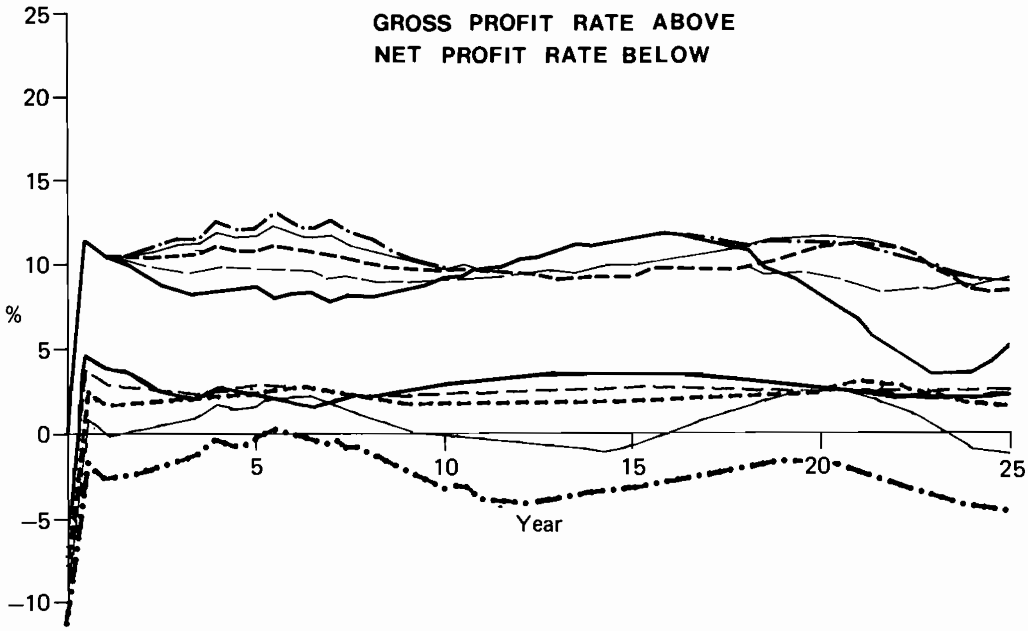


Figure 5 Profit Rates with Unlimited External Financing **(Run 2)**

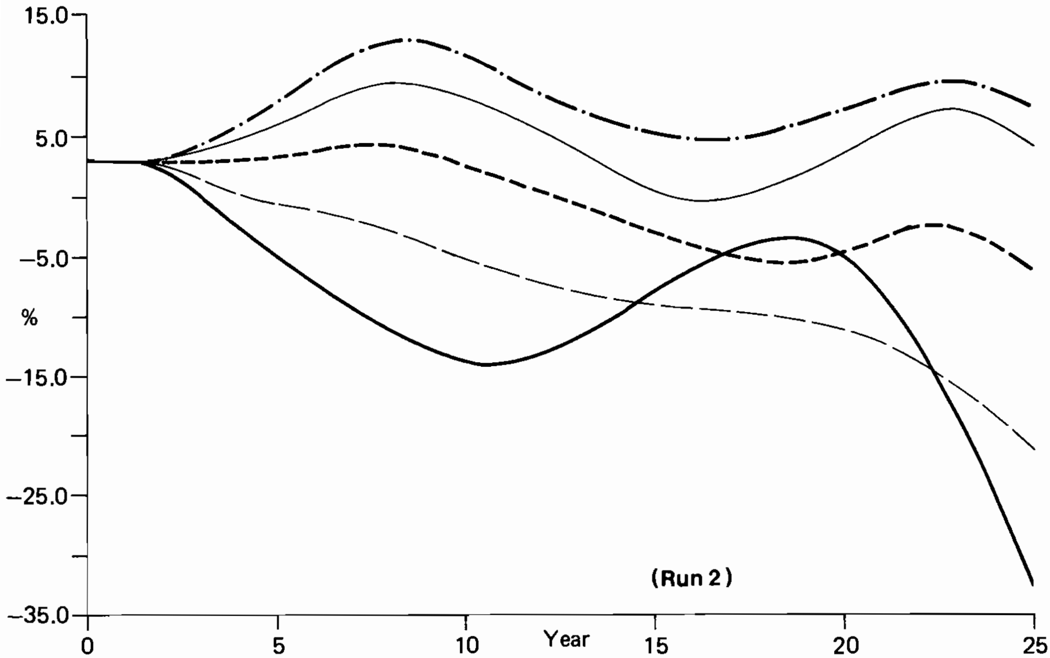


Figure 6 Desired (and Realized) Expansion Rates, Unlimited External Financing Case

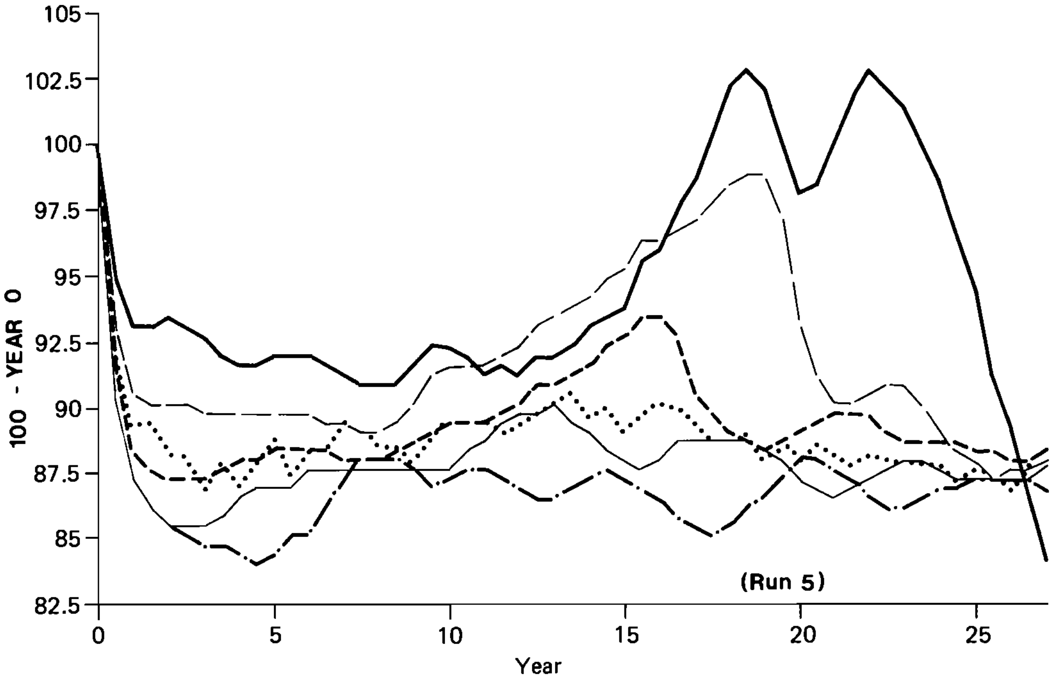


Figure 7 Price Indices and Average Price Level (Dotted) for Internal Financing Case

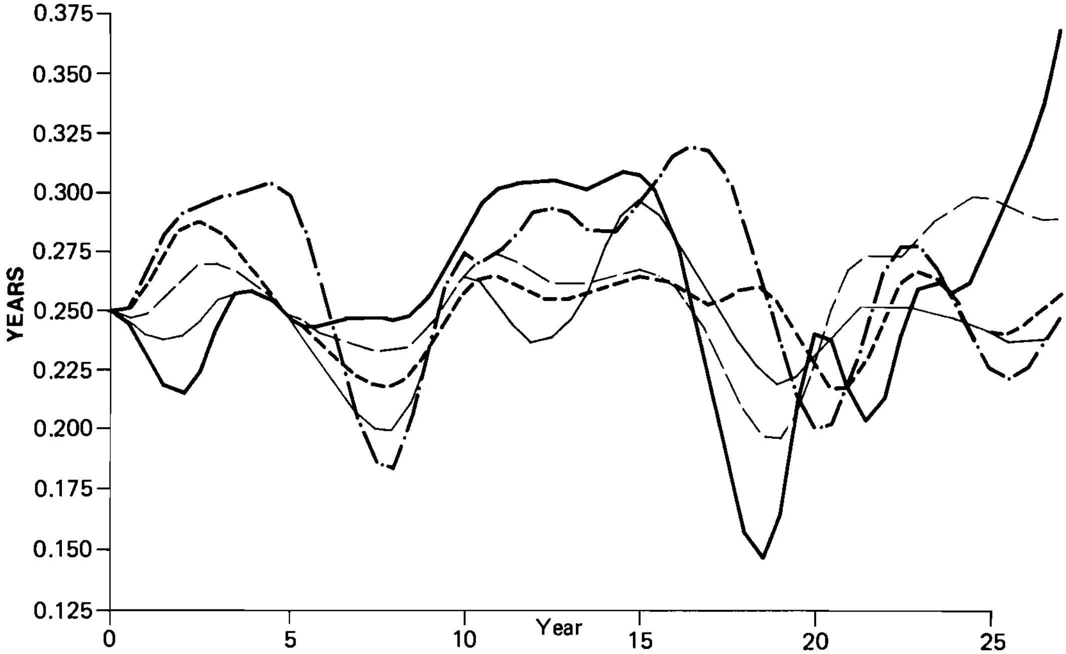


Figure 8 Delivery Delays for Internal Financing Case (Run 5)

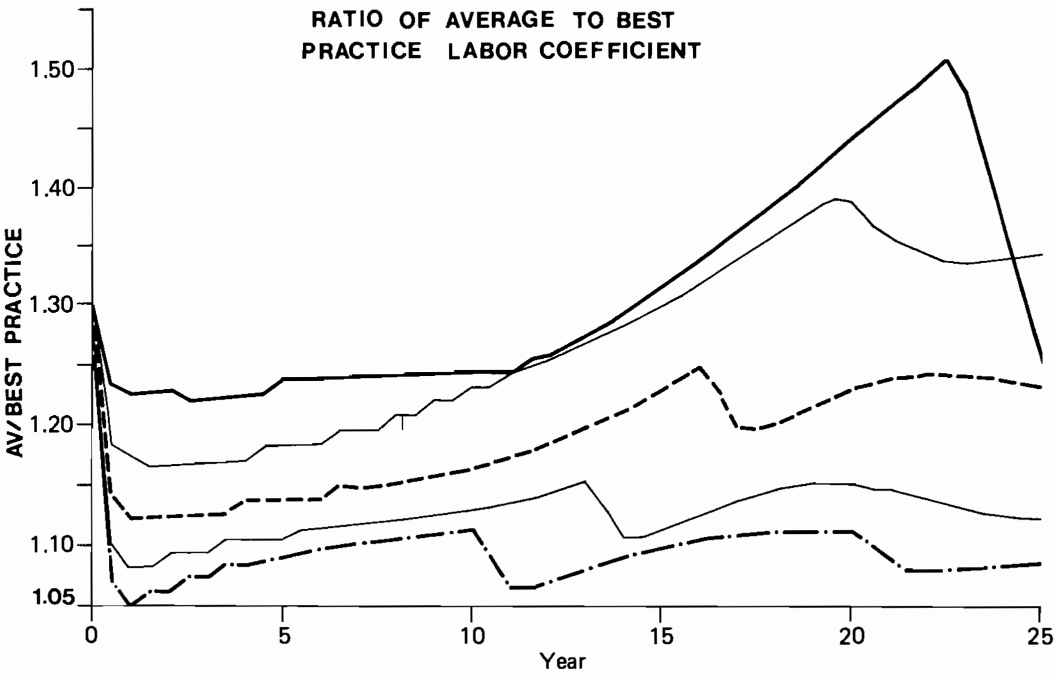


Figure 9 Unlimited External Financing Case **(Run 2)**

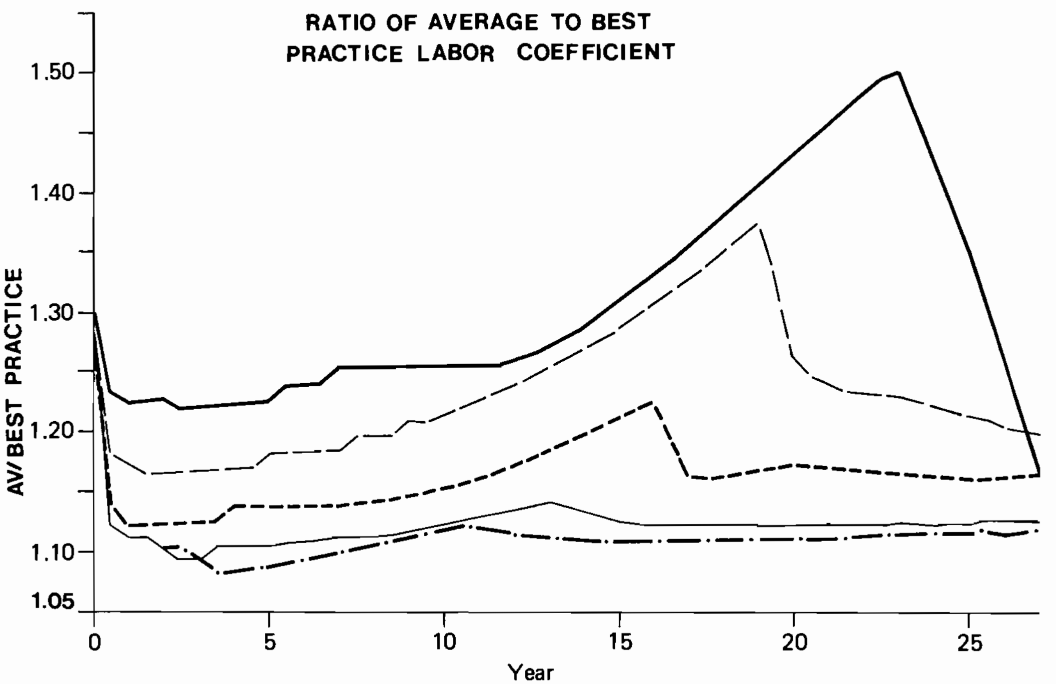


Figure 10 Internal Financing Case **(Run 5)**

differ only with respect to the financing mode. Although firm 1 aims at a 10 year replacement policy and firm 2 a 13 year one, because firm 2 can devote more financing to expansion in the second run, the difference between their ratios narrows appreciably.

If we now return to the behavior of market shares under internal financing (Figure 4), one perceives that a reversal has taken place between firms 1 and 2 for market leadership. This is due to the fact that, although firm 1 has a cost advantage due to its younger capital stock, its failure to expand sufficiently due to a lack of internal finance gives it a delivery delay disadvantage. Thus we see that the combination of the ability to internally finance investment, and the relationship between capacity and delivery delay, both enter crucially into the determination of optimum replacement policy.

7. PROSPECTS FOR FURTHER RESEARCH

The above model is the first step in a research program aimed at modelling structural change in historical time as an ongoing process, critically dependent on microeconomic diversity and expectational dynamics deriving from fundamental uncertainty. The next stage is to allow for true choice of best-practice technique and learning-by-doing effects associated with a given technological trajectory (see Dosi, Orsenigo and Silverberg, 1986). It is also intended to close the model in one and two sectoral versions along the lines of a previous paper (Silverberg 1984) by incorporating an unemployment-nominal wage growth relation.The possibility of empirical analysis with this kind of non-steady-state vintage model (however, at an aggregate level without self-organizational dynamics) has already been demonstrated by Soete and Dosi (1983) and McIntosh (1986). It is hoped that in the near future appropriate methods and data can be found to permit empirical analysis at the disaggregated - market share and strategic decision making - level as well.

REFERENCES

Axelrod, R., The Evolution of Cooperation, Basic Books, New York, 1984.

Bains, J.S., "The Relation of the Economic Life of Equipment to Reinvestment Cycles", Review of Economic Studies, 21, 1939, pp. 79-88.

Clark, J.A., "A Model of Embodied Technical Change and Employment", Technological Forecasting and Social Change, 16, 1980, pp. 47-65.

Cyert, R.M., Debrost, M.H., and Holt, C.A., "Capital Allocation within a Firm", Behavioral Science, 24, 1979, pp. 287-95.

Dosi, G., Orsenigo, L., and Silverberg, G., "Innovation, Diversity, and Diffusion: A Self-Organisation Model", Paper Presented at the International Conference on Innovation Diffusion, Venice, 17-21 March 1986.

Ebeling, W., and Feistel, R., Physik der Selbstorganisation und Evolution, Akademie-Verlag, Berlin (East), 1982.

Eigen, M., "Self-Organisation of Matter and the Evolution of Biological Macro-Molecules", Naturwissenschaften, 58, 1971, pp. 465-522.

Eisner, R., Determinants of Capital Expenditures. An Interview Study, University of Illinois Press, Urbana, Ill., 1956.

Eiteman, W.J., Price Determination in Oligopolistic and Monopolistic Situations, University of Michigan Press, Ann Arbor, 1960.

Haken, H., Synergetics, 3rd ed., Springer-Verlag, New York-Heidelberg-Berlin, 1983.

Kaldor, N., "Gemeinsamkeiten under Unterschiede in den Theorien von Keynes, Kalecki and Rüstow", IFO-Studien, 29, 1983, pp. 1-10.

Kaldor, N., and Mirrlees, J.A., "A New Model of Economic Growth", Review of Economic Studies, 29, 1962, pp. 174-92.

Kalecki, M., Theory of Economic Dynamics, Monthly Review Press, New York, 1965.

McIntosh, J., "Economic Growth and Technical Change in Britain 1950-1978", European Economic Review, 30, 1986, pp. 117-28.

Nelson, R.R., and Winter, S.G., An Evolutionary Theory of Economic Change, Belknap Press, Cambridge, Ma., 1982.

Neumann, J.V., and Morgenstern, O., The Theory of Games and Economic Behavior, Princeton University Press, Princeton, 1947.

Nicolis, G., and Prigogine, I., Self-Organization in Non-Equilibrium Systems, Wiley, New York, 1977,

Prigogine, I., "Order through Fluctuation: Self-Organization and Social System", in Jantsch, E., and Waddington, C.H., (eds.), Evolution and Consciousness, Addison-Wesley, Ma., 1976.

Rosenberg, N., "On Technological Expectations", Economic Journal, 86, 1976, pp. 523-35.

Rüstow, H-J., Theorie der Vollbeschäftigung in der freien Markwirtschaft, J.B.C. Mohr, Tubingen, (Paul Siebeck), 1951.

Rüstow, H-J., Neue Wege zur Vollbeschäftigung. Das Versagen der ökonomishcen Theorie, Duncker & Humblot, Berlin, 1984.

Salter, W., Productivity and Technical Change, Cambridge University Press, Cambridge, 1962.

Schwartz, J., Theory of Money, Gordon and Breach, New York, 1965.

Silverberg, G., "Embodied Technical Progress in a Dynamic Economic Model. The Self-Organization Paradigm", in Goodwin, R.M., Kruger, M., and Vercelli, A., (eds.), Nonlinear Models of Fluctuating Growth, Springer-Verlag, Berlin-Heidelberg-New York, 1984.

Simon, H.A., "Theories of Decision Making in Economics", American Economic Review, 49, 1959, pp. 253-83.

Smale, S., The Mathematics of Time. Essay on Dynamical Systems, Economic Processes,

and Related Topics, Springer-Verlag, New York-Heidelberg-Berlin, 1980.
Smith, V., Investment and Production, Harvard University Press, Cambridge, Ma.,1961.
Soete, L., and Dosi, G., Technology and Employment in the Electronics Industry, Francis Pinter, London, 1983.
Solow, R., "Investment and Technical Progess", in Arrow, K.J., Karlin S., and Suppes, P., (eds.), Mathematical Methods in the Social Science, Stanford University Press, Standford, 1960.
Terborgh, G.,Dynamic Equipment Policy, MacGraw-Hill, New York, 1949.
Williamson, O.E., "A Dynamic Theory of Interfirm Behavior", Quarterly Journal of Economics, 79, 1965, 579/607.

Chapter 7

Technological Vintages and Substitution Processes

B. JOHANSSON

1. INTRODUCTION

1.1 Structural Adjustments and Invariances in Industrial Sectors

An important aspect of Schumpeter's (1934) vision of the process of economic development is that early initiators receive a premium for being forerunners. This "extra profit" to innovators has the form of a temporary monopoly, based on specific knowledge that they do not share with their competitors - or only share with a few of them.

Assuming that such innovations occur continuously over time, or in time intervals with varying periods of imitation, one should expect productivity and profits to be unevenly distributed over the firms in a sector. High profit and productivity levels would then be associated with the leaders and the lower levels with the followers.

In Sweden and Norway, information is available about profit and productivity distributions over individual production units (establishments) with a fine sector specification within the manufacturing industry. This type of data indeed confirms that the economic reward to production units are "Schumpeter distributed", and that such distributions have a characteristic form (see Johansson and Marksjö, 1984; Johansson and Strömqvist, 1981).

For each given sector, not only the general form, but also the specific parameters of such distributions remain time invariant over long intervals. This property, which has been examined and documented in detail for the period 1968-1984, is the result of ongoing dynamic processes. The current essay outlines some preliminary models in an attempt to explain the observed patterns. The focus is on the exit and entry of production techniques and products in a sector. An earlier attempt to model these processes can be found in Johansson and Holmberg (1982).

Innovations and utilization of technological knowledge in a sector appear in two distinct forms; each establishment may renew its production technique, but it may also adjust its old technique and develop new products. The effects of introducing new products are analyzed within a framework related to Lancaster (1971). The substitution between old and new products is compared and combined with the dynamic substitution between old and new production techniques. In summary, the study deals with dynamic processes which generate specific distributions of process and product vintages associated with the observed profit and productivity distributions.

1.2 Outline of the Study

In section 2, the above-mentioned form of observed profit distributions is described and basic model elements are introduced. The representation of process (technique) and product vintages are formally introduced.

Section 3 deals with the entry and exit of process vintages. Different assumptions about the entry and exit dynamics generate different forms of the profit distribution in a sector. This section concentrates on the assessment of such distribution effects of small variations in assumptions.

Section 4 describes how product change with logistic substitution processes may help to explain the steepness of empirically observed profit distributions. In the absence of product evolution, technical change will generate profit distributions which are too flat.

The paper ends with a summary describing the research directions which are pointed out in this tentative investigation.

2. PRODUCT AND PROCESS VINTAGES

2.1 The Form of Productivity and Profit Distributions

It is an intriguing fact that the distribution of productivity among establishments in an industrial sector displays a shape which is approximately invariant over periods of considerable length. Moreover, the general form of such distributions may be fitted with rather good precision to a curve of logistic type.

To make the above statements more concrete, consider a sector with individual production units $i = 1,2,....$ Let p_i be the price of unit i's production, x_i, and let c_i be its unit cost of production. Then we may define

$$\pi_i = p_i - c_i \tag{2.1}$$

$$\mu_i = p_i x_i / L_i \tag{2.2}$$

where π denotes gross profit per unit output and μ the productivity, given that L represents an appropriate factor of production, e.g., labour input.

Imagine now that we arrange all units in descending order according to the size of π_i and μ_i, respectively. Let, in each case, $\pi_{i+1} < \pi_i$ and $\mu_{i+1} < \mu_i$.We may then form the following two cumulative distributions $z(\pi_i)$ and $z(\mu_i)$:

$$z(\pi_k) = \sum_{i \leqq k} x_i/x \tag{2.3}$$

$$z(\mu_k) = \sum_{i \leqq k} x_i/x \tag{2.4}$$

where $x = \Sigma_i x_i$ denotes total production in the sector. The distribution in (2.3) is illustrated in Figure 2.1. The associated distribution in (2.4) will have a similar form and will be a scale transformation of the first. Under modest assumptions, the π-ranking of units will be equivalent with the μ-ranking, and then we may write $\pi_i = \alpha \mu_i + \epsilon_i$, where ϵ is a randomly distributed error term.

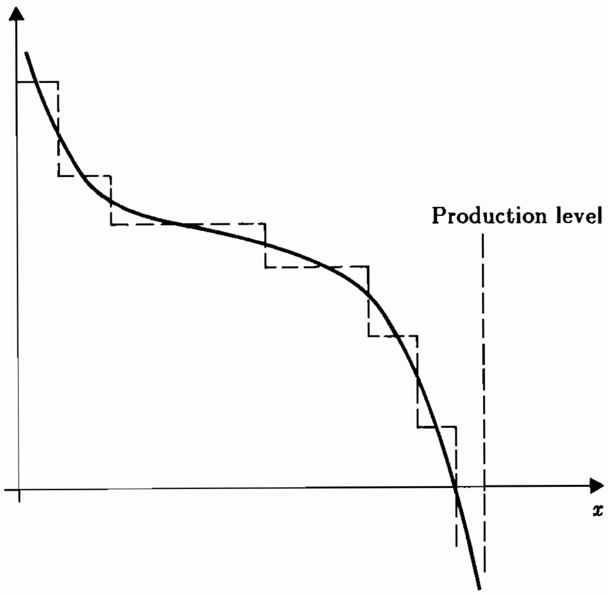


Figure 2.1 Logistic distribution of gross profits over production capacity

The invariance of productivity distributions has been reported in Johansson and Marksjö (1984). In that case the productivity was defined as

$$\mu_i = (p_i x_i - H_i)/L_i$$

where H_i denotes non-labour variable input costs and L_i the number of employees in production unit i. Over a range of sectors and economic regions, the following relationship was shown to fit observed data in a statistically acceptable way:

$$\ln z = a_0 - a_1 \mu + a_2 t \quad (2.5)$$

where $a_2 > 0$ translates the μ-curve as time, t, increases. The coefficient $a_1 > 0$ describes how the logarithm of the share of total employment with $\mu > \bar{\mu}$ decreases as $\bar{\mu}$ increases.

The result in (2.5) illustrates a considerable structural invariance. This paper aims to describe the processes which bring about this property. Another observation in Johansson and Marksjö (1984) is that the variation in π_i/p_i is small between pairs of years t and t+1, and with fixed price evaluations, this variation is even smaller. These observations form a basis for treating the profitability and productivity differentials as reflections of a vintage distribution. We also attempt to establish that such differentials in an industry represent a combination of (i) production technique vintages and (ii) product vintages.

2.2 Representation of Process Vintages

In the sequel we will assume that if a vintage index is attached to each production unit in a sector, this index will represent a combination of the unit's process vintage and product vintage. In this subsection, we assume that all units produce a similar type of product and concentrate on production process vintages or differentials as regards the production techniques applied in the sector.

Given the assumption above, the relation between a production unit's profit, π, and its process vintage τ, may be schematically described in the following way:

$$\pi(\tau) = p - c(\tau) \qquad (2.6)$$
$$c(\tau) = \tau\bar{c}$$

where $\bar{c}$ is a constant, p denotes unit price and $c(\tau)$ unit costs given technique τ. We may assume that the best practice technique at time t, $\tau^*(t)$, is improved over time so that $\tau^*(t) > \tau^*(t+\Delta)$. If all investments are made with the current best practice at each time, the τ-index will indeed also function as a vintage index.

A production unit applying technique τ also has a production capacity $\bar{x}(\tau)$, which constrains current output so that $x(\tau) \leqq \bar{x}(\tau)$. Assume now that demand, Y, is given as a function of time, t, and price

$$Y = Y(p,t) \qquad (2.7)$$

and define excess demand, y, as

$$y(t) = Y(p,t) - \bar{x}(t) \qquad (2.8)$$

$\bar{x}(t) = \Sigma_\tau \bar{x}(\tau,t)$.

One may also use (2.7) and (2.8) in respecifications of (2.6) to obtain the following expressions (for each given point in time):

$$\pi(\tau,y) = p(y) - c(\tau,y) \qquad (2.6')$$
$$c(\tau,y) = \tau\bar{c}(y).$$

At a later stage y is used to generate dynamic adjustments.

2.3 Product Attributes and Product Vintages

In this subsection, we present a framework within which we may distinguish between old and new products, and the way in which new product vintages substitute older vintages.

First, we introduce the concept of a product group I. In this group we identify a product type i by the amount θ_j^i of attribute $j = 1,...,m$ contained in one unit of the product. Product i is then characterized by the vector $\theta^i = (\theta_1^i,...,\theta_m^i) \in R_I^m$. Every product in group I satisfies $\theta_j^i \geqq 0$ for all j and $\theta_k^i > 0$ for at least one k.

Following Lancaster (1971), we assume that the demand for products is derived from customers' demand for attributes. We identify customer groups $g = 1,2,...,$ and for each group we introduce a mapping B^g such that

$$B^g: G_I \rightarrow R_I^m \qquad (2.9)$$

where G_I is the space of product vectors $x_I = \{x_j\} \in G_I$, and where B^g describes how customer group g may combine different products $j \in I$ to obtain a vector $\theta = B(x_I)$ of attributes.

We also assume that each customer group has preferences, defined by a preference function u^g:

$$u^g: R_I^m \to R \tag{2.10}$$

At each point in time we may identify group g's budget m_I^g that constrains its purchases of products $j \in I$. Demand may then be derived from the following decision problem

$$\max\ u^g(B^g(x_I))$$
$$\text{subject to} \tag{2.11}$$
$$m_I^g \geqq \sum_{j \in I} p_j x_j$$

to obtain a demand function for each i

$$Y_i(p,m_I) = \Sigma_g\ Y_i^g(p,m_I^g) \tag{2.12}$$

where $p = \{p_j\}$, $j \in I$ and $m = \{m^g\}$.

According to (2.12) the introduction of a new product type i may cause a process such that $\dot{Y}_i > 0$ as the result of $\dot{p}_i < \dot{p}_j$ for all $j \neq i$, and $\dot{m} > 0$. Naturally, such a change implies that other products are substituted partially or completely, as analyzed in section 4. There we also show that if a new product k satisfies $\theta_k = \lambda \theta_i$ with $\lambda > 1$, the product vintages k and i may be modelled as two different process vintages.

3. ENTRY AND EXIT OF PROCESS VINTAGES

3.1 Substitution Between Production Techniques

The change of a sector's vintage structure is generated by (i) entry and (ii) exit of production techniques. The first process is caused by two types of investments. New capacities embodying a new technique may be created, and existing capacities may be improved, by changing the technique from τ^0 to τ. Let $q(\tau^0\text{-}\tau;\ t)$ denote the annual cost per unit capacity, associated with an investment that changes the technique from τ^0 to τ. For a new production unit, the same cost is denoted by $q(\tau;\ t)$. With this notation a profit-motivated investment criterion can be formulated as follows:

$$\pi(\tau^*) - \pi(\tau^0) = \pi(\tau^*,\tau^0) \geqq q(\tau^0-\tau^*) \geqq 0 \tag{3.1}$$

where $\tau^0 = \tau(t_0)$ and $\tau^* = \tau(t_*)$ represents vintages associated with time t_0 and t_*, respectively. In fact, (3.1) should also apply for $\pi(\tau,\tau^0)$ where $\tau^* > \tau > \tau^0$, a result which obtains if capacity $\Delta x(\tau^*)$ is added to an existing capacity $x(\tau^0)$.

In established vintage analysis (Salter, 1960; Johansen, 1972 etc.), one assumes that old capacities with technique τ^0 are removed at a time t such that $\pi(\tau^0,t) = 0$. Following observations reported in Strömquist (1983), we assume here that capacity removal is a gradual phenomenon with an increased removal frequency as process techniques become economically obsolete. This exit process is described by an exit function ϵ such that

$$\epsilon(t) = \epsilon(\pi(\tau,t)) \geqq 0$$
$$\partial\epsilon/\partial\pi < 0. \tag{3.2}$$

Given these preliminaries, let us study a case in which the sector has only one type of product, in which the price path $p(t)$ is given, and $y \geqq 0$ at all points in time. Employing these assumptions we may express ϵ as functions of τ and t. Next, consider a division of the total production capacity into shares, $\sigma_k(0)$ at time $t = 0$, such that $\Sigma\sigma_k(0) = 1$; at time t the technique is denoted by $\tau_k(t)$.

The removal in technique class k is $\epsilon_k x_k = \epsilon(\tau_k)x_k$. Suppose, as an initial example, that capacity with the new technique τ^* is introduced into each class as a proportion of expected change in market-clearing capacity, expressed by

$$\dot{v} = \alpha y,\ 0 \leqq \alpha \leqq 1, \tag{3.3}$$

if $y > 0$ and $\dot{v} = 0$ otherwise. The capacity change in x_k is then described by

$$\dot{x}_k = -\epsilon(\tau_k)x_k + \sigma_k\dot{v} \tag{3.4}$$

for $\sigma_k = \sigma_k(0)$. When y and ϵ_k are approximately constant, (3.4) yields $x_k(t) \approx x_k(0)\ \exp\{-\epsilon_k t\} + (\alpha y \sigma_k/\epsilon_k)(1-\exp\{-\epsilon_k t\})$ so that $x(t) = \Sigma\ x_k(t)$ approaches $\alpha y/\epsilon$ and ϵ_k approaches $\epsilon = \epsilon(\tau^*)$.

Naturally, the process in (3.3)-(3.4) generates a change such that $x_k(t)/x(t)$ approaches $\sigma_k(0)$ and $\tau_k(t) = \tau^*$ for all k as t expands. The result is a flat technique distribution.

The above solution represents an extreme case which corresponds to a technically stagnating sector. However, the cases observed in (2.5) imply an unchanged relative position of the various technique classes k. Such a change process becomes closer if (3.4) is exchanged for

$$\begin{aligned} \dot{x}^0_k &= -\epsilon(\tau_k)x^0_k(t) \\ \dot{x}^*_k &= -\ x^0_k \\ x_k(t) &= x^0_k(t) + x^*_k(t) \end{aligned} \tag{3.5}$$

where $x^0_k(t)$ and $x^*_k(t)$ denotes the capacities within class k employing technique $\tau^0_k = \tau_k(0)$ and τ^*, respectively. For class $k = 1,...,N-1$ (classes established before t=0) $\tau_k(t)$ will gradually approach τ^*; $\tau_k(t) = \tau^* + (\tau^0_k - \tau^*)\ \exp\ \{-\epsilon_k t\}$, provided that $\epsilon(\tau_k) = \epsilon_k$.

At t = 0 we let N denote the new class which is introduced according to the following process:

$$\dot{x}^*_N = \alpha(\tau^*)\ \max\ \{0,y\} - \epsilon(\tau^*)x^*_N \tag{3.6}$$

$$\alpha(\tau^*) \begin{cases} > 0 \text{ if } \pi(\tau^*,y) > q(\tau^*) \\ = 0 \text{ otherwise} \end{cases}$$

where $q(\tau^*)$ denotes the current capital cost associated with investment in capacity of vintage τ^*.

If we assume $y \geqq 0$ in (3.3), we need only add $\epsilon(\tau^*,t) = 0$ to obtain

$$x^*_N(T) = \int_0^T \alpha(\tau^*,t)y(t)dt. \tag{3.7}$$

Where $\alpha(\tau^*,t) = \alpha$ and y is constant, this yields a smooth increase of $x^*_N/x = x^*_N(t) \;/\; (x^*_N(t) + x(0))$. Since each other class substitutes technique τ_k and replaces it with τ^*, the long term result is a flat profit curve.

However, (3.5) and (3.6) still do not produce a profit pattern of the kind described in (2.5) and Figure 2.1. In order to model a process which may generate such a structure, it is also necessary to consider technical change as an ongoing process in the sense that the best practice technique evolves as time goes by. A simple way of describing this is to assume that the best practice is improved at successive points in time in such a way that we obtain $\tau^*(0) > \tau^*(t_a) > \tau^*(t_b)$ etc. at $0 < t_a < t_b$. The τ^*-values may be selected as points on the path $\tau^*(t) = \tau^{**} + \Phi \exp\{-\lambda t\}$, $\lambda > 0$, for the different dates of technical innovation, where τ^{**} denotes a lower bound for the best practice, given by physical constraints on the given production process.

First, we denote the best practice class at time t by $N(t)$ and apply formula (3.6) to the development of $x^*_{N(t)}$. Second, we change (3.5) to read

$$\begin{aligned}
&\dot{x}_k = -\epsilon(\tau_k)x_k(t) \\
&\dot{x}^*_k = \theta_k\alpha(\tau^*)x^*_N(\tau^*(t-\lambda_N),t-\lambda_N)\sigma_k \\
&0 \leqq \theta_k < 1 \;;\; \lambda_N \geqq 0 \\
&\sigma_k = x_k(t)/\Sigma_k\, x_k(t).
\end{aligned} \tag{3.5'}$$

The formulation in (3.5') reflects the assumption that technical renewal in established production units occurs as imitation of the technique applied in units of vintage $N(t-\lambda_N)$. The parameter λ_N then has the nature of an imitation lag and θ_k

represents the intensity of technical adoption.

At time $t \geqq t_b$, the development of vintage $N(t_a)$, $t_a < t_b$ is switched over from process (3.6) to (3.5'). Combining (3.6) and (3.5'), we may consider variations in the steepness of the distribution curve in Figure 2.1. Let $z(\bar{\tau})$ denote the proportion of capacities employing techniques $\tau \leqq \bar{\tau}$. We may measure steepness as the size of $a_1 > 0$ in the equation

$$\log z(\bar{\tau}) = a_0 - a_1 \pi(\bar{\tau}) \tag{3.8}$$

which approximates the actual pattern, or simply as the ratio

$$\sum_{\tau \leqq \bar{\tau}} \pi(\tau) x(\tau) / \sum_{\tau > \bar{\tau}} \pi(\tau) x(\tau). \tag{3.9}$$

Remark 1. Assume that the change pattern is generated by (3.6) together with (3.5'), and that new techniques are initiated at successive dates $0 < t_a < t_b < \ldots.$ Then the following conditions obviously cause increased steepness, as measured by (3.8) or (3.9), of the profit distribution curve, provided that $\alpha(\tau^*) > 0$:

(i) A high level of excess demand, $y > 0$, and a low or zero rate of exit, $\epsilon(\tau^*(t))$, for the best practice technique, $\tau^*(t)$;

(ii) a low imitation intensity θ_k and a high imitation lag λ_k together with a low rate of exit $\epsilon(\tau_k)$ for $k > \tau^*(t)$.

We may also conclude that a flatter, less steep, distribution curve is generated if the conditions (i)-(ii) in the remark are reversed. This leads to the following remark which characterizes the ultimate effects of technical change.

Remark 2. Consider the three different change processes described by (3.4); (3.5) and (3.6); (3.5'), (3.6) and $\tau^*(t) = \tau^{**} + \Phi \exp\{-\lambda t\}$. In the long term all three processes lead to flat profit distributions.

Hence, from a technical change or vintage renewal perspective, we may conclude that a gradual reduction of steepness of the profit distribution (for an industry characterized by a given product type) reflects that the industry is ageing technologically.

3.2 Adjustment Mechanisms of Technical Change

The introduction of new vintages is stimulated by a positive excess demand, $y = Y - x$, as described in (3.6). Excess demand is increased over time if $Y(p,t)$ grows. Such an increase may be exogenously driven by time, e.g., by the emergence of new customer groups. In particular, we have assumed that $Y' = \partial Y / \partial p < 0$, which means that a gradual change $\dot{p} < 0$ will generate a continuing growth in demand. Setting $\dot{Y} = (\partial Y / \partial p)\dot{p}$ we can see that $\dot{p} < 0$ stimulates a positive excess demand, since $\dot{y} = \dot{Y} - \dot{x}$.
First consider the long term process of price changes. Observe first that we have assumed that $\tau^*(t) = \tau^{**} + \Phi \exp\{-\lambda t\}$ which yields

$$\dot{\tau}^* = -\lambda\, \Phi \exp\{-\lambda t\}. \tag{3.10}$$

Observe next, that from (2.6) we have $\pi(p,\tau) = p - \tau\bar{c}$. Moreover, from (3.1) we have that technique τ^* is introduced only if $\pi^* - q(\tau^*) = \bar{q} \geqq 0$. Hence, observing that $\dot{\pi}^* = \dot{p} + \lambda\Phi \exp\{-\lambda t\}\, \bar{c}$ we may formulate the following conclusion:

<u>Remark 3</u>. Given (2.6), (3.1) and (3.10), the following price adjustment satisfies the investment criterion for the successively available best practice:

$$\dot{p} = \dot{q} - \lambda\Phi \exp\{-\lambda t)\, \bar{c}.$$

For $\dot{q} \approx 0$, the price will be reduced at a gradually retarded pace until technical change terminates.

From (3.6) and (3.5') we may approximate the exit process with $E(p)$ such that

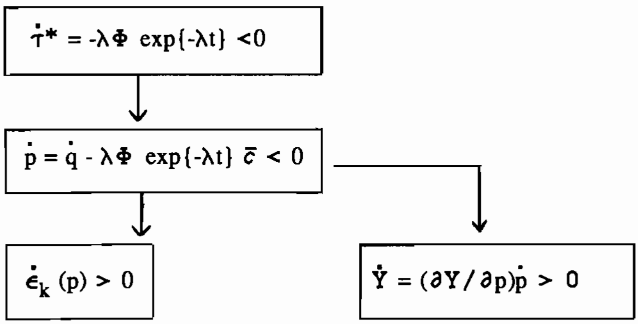
$$E(p) = \sum_{k=1}^{N(t)} \epsilon_k(p) x_k \tag{3.11}$$

where $\epsilon_k(p) = \epsilon(\pi(p,\tau_k))$. From (2.6) and (3.2) we may derive $\partial\epsilon/\partial p < 0$ so that where $\epsilon_k(p) = \epsilon(\pi(p,\tau_k))$. From (2.6) and (3.2) we may derive $\partial\epsilon/\partial p < 0$ so that

$$\partial E/\partial p < 0. \tag{3.12}$$

<u>Remark 4</u>. The price adjustment in Remark 3 satisfies the investment criterion and creates investment opportunities by intensifying the exit process due to (3.12). The resulting change process is illustrated in Figure 3.1.

We may observe that the statement in Remark 4, and the corresponding description in Figure 3.1, constitute open statements, i.e., the system is not closed. Without entering any penetrating analysis of the dynamics, we outline the nature of the "short term" price adjustments that must complete the "long term" change pattern, which is described in Remark 3.



<u>Figure 3.1</u> Technology-induced price adjustments

Let us write max $\{0,y\} = F_N(y)$, $\dot{p} = f(y)$, and assume tentatively that

$$f(y) \begin{cases} > 0 \text{ if } y > \bar{y} \geqq 0 \\ = 0 \text{ if } y = \bar{y} \\ < 0 \text{ otherwise} \end{cases} \tag{3.13}$$

$$f' = \partial f/\partial y > 0.$$

Returning to (2.8) we observe that $y(p,t) = Y(p,t) - x(t)$. Using (3.6), (3.5') and (3.11) we may write (for zero imitation lag)

$$\dot{y} = Y'\dot{p} - \alpha F_N(y) - \sum_{k<N} \theta_k \alpha x^*_N \ \sigma_k + E(p) \tag{3.14}$$

where $Y' = \partial Y/\partial p < 0$ and $\partial Y/\partial t = 0$. The condition $\dot{y} = 0$ may be selected as one kind of equilibrium. We may also consider a situation with both $\dot{y} = 0$ and $\dot{p} = 0$ as an equilibrium. The strongest equilibrium property obtains when we combine $\dot{y} = 0$ with the technology-induced price adjustment in Remark 3.

Remark 5. At any given state, let $H = \alpha F_N(y) + x^*_N \Sigma\theta_k\alpha\sigma_k$. Then it follows from (3.14) that the equilibrium condition $\dot{y} = 0$ is generated by a price adjustment process

$$\dot{p} = [H - E(p)] / Y'.$$

We may conclude that this is a more complicated adjustment mechanism than the one in (3.13).

To verify whether the last statement is true, we may note that the equilibrium price adjustment in the remark is $f:(y,x_1,\dots,x_{N-1},x^*_N) \to \dot{p}$ which reacts not only to y but to the entire vintage pattern.

Remark 6. From (3.14) we may conclude that a situation with $\dot{y} = \dot{p} = 0$ obtains if

(i) $y = 0, \partial Y/\partial t = 0$, $\epsilon_N(p) = 0$, and if each θ_k is a function of p such that $\theta_k(p) = \epsilon_k(p)/\alpha\sigma^*_N$, where σ^*_N is constant, or

(ii) $y = \bar{y} > 0, dY/dt = \alpha y$, and if $\theta_k(p,x^*_N) = [\epsilon_k(p)/\alpha - y/x]\ \sigma^*_N$.

Part (i) in the remark means that capacity which is removed from old vintages is also replaced by exactly the same amount of best practice. Observe first that $\dot{p} = 0$ if $(H\text{-}E(p)) = 0$. Next, we observe that $\sigma_k x^*_N = \sigma^*_N x_k$. From $y = 0$ and $\epsilon_N(p) = 0$ it follows that $\dot{x}^*_N = \alpha y = 0$. Hence, $H - E(p) = 0$ if for each k, $\theta_k(p)\alpha\sigma^*_N = \epsilon_k(p)$ which implies that each $\dot{x}_k$=0. This together with $\partial Y/\partial t = 0$, $\dot{x}^*_N = \dot{p} = 0$ implies that $\sigma^*_N = x^*_N/x$ is constant.

Part (ii) makes use of the fact that if $\theta_k(p,x^*_N)\sigma^*_N x_k\alpha + \alpha\bar{y}x_k/x - \epsilon_k(p)x_k = 0$ for each k, then H - E(p) in Remark 5 will be zero. Hence, $\dot{y} = \dot{p} = 0$, and y may remain positive to generate a growth in x^*_N.

The two different conditions in Remark 6 both require that the imitation intensity, $\theta_k(p)$, is balanced against the rate of exit, $\epsilon_k(p)$. With condition (ii) it also follows that a steady growth in demand, together with a slow exit rate for old vintages, will correspond to a low imitation intensity, and all this will induce an increased steepness of the profit distribution.

4. ENTRY AND EXIT OF PRODUCTS

4.1 Product Substitution

Subsection 2.3 constitutes the starting-point of the following analysis describing how a new product may gradually acquire an increased market share.

First, consider the mapping B^g in (2.9). In order to obtain a simple analytical structure, we assume here that $B^g = B$ is the same for all customer groups g, and that B is an irreducible matrix built up by a non-negative attribute vector θ^i for each product type i. Hence, we may specify (2.9) in the following form

$$z = Bx_I \qquad (2.9')$$

where $z \in R^m{}_I$ is the attribute vector a customer obtains by using (purchasing) given amounts of each product as specified in the product vector x_I. Next, assume that the preference functions are well behaved: u^g is continuous, strictly quasi-concave and differentiable with all first order derivatives positive. Given this assumption, together with assumption about B in (2.9'), we may conclude that the demand functions in (2.12) are upper semi-continuous (Berge, 1959). In addition, each Y_i-function is continuous if there is no product $i \in I$ whose attribute vector is a linear combination of some other attribute vectors in the product group (see, e.g. Johansson, 1978). In this case, we can describe the solutions to (2.11) as

$$x_i^g = \mu_i^g(p)m^g/p_i$$
$$x_i = \Sigma_g x_i^g \qquad (4.1)$$
$$0 \leqq \mu_i^g \leqq 1$$

where μ_i^g is a continuous function of prices and expresses the share of group g's budget (for market I), which is allocated to the purchase of product i (compare with Batten and Johansson, 1985). Hence, μ_i-values represent market shares. Summing over customer groups (industrial sectors, households etc.) we may write $\mu_i(p)m/p_i = \Sigma_g \mu_i^g(p)m^g/p_i$ for $m = \Sigma m^g$ and describe the change in μ_i as

$$\dot{\mu} = k_i(t)\mu_i(1 - \mu_i) \qquad (4.2)$$

where k_i may vary over time. If k_i remains constant, (4.2) equals Verhulst-Pearl's equation which describes a logistic growth path of μ_i (or logistic decline pattern). The process in (4.2) is generated by changes in attribute vectors and/or in relative prices. If these changes are "steady over time", k_i may remain constant.

The conditions given by (2.11), (2.9') and the associated assumptions about B and u^g imply that each μ_i^g increases as p_i is reduced relative to other prices in the commodity group I. With this property we may consider the following description of the development path of μ_i when the market consists of only two products, 1 and 2:

$$\ln(\mu_1/\mu_2) = a_0 + a_1 t + b(p_1/p_2) \qquad (4.3)$$

or $\ln x_1/x_2 = a_0 + a_1 t + b(p_1/p_2) - \ln(p_1/p_2)$, where a_0 is a constant, $a_1 \geqq 0$, $b \leqq 0$, and t denotes time. Such a function describes the introduction of new products and usually performs better than just $\ln \mu_1/\mu_2 = a_0 + a_1 t$, which is the standard introduction equation. If price changes are the only driving force, we have $a_1 = 0$. Formula (4.3) corresponds to a case in which k_i in (4.2) satisfies $k_1 = k_1(p_1,p_2)$.

4.2 Properties of the Substitution Process

The vector $\theta^i = (\theta^i{}_1,\ldots,\theta^i{}_m)$ describes the amount of attributes obtained from one unit of product i. From (4.1) we can see that customer group 1 obtains the following attributes by combining products

$$z(p)^g = \Sigma_i \mu^g{}_i(p)\theta^i m^g/p_i, \tag{4.4}$$

and $z^i(p)^g = \theta^i m^g/p_i$ is the attribute amount that obtains if the whole budget m^g is allocated to product i. Given these elaborations, we may illustrate the connection between product and process innovation.

> Remark 7. Let $\theta > 0$ and assume that products 1 and 2 belong to the same product group ($\theta^1 \epsilon R^m{}_I$ and $\theta^2 \epsilon R^m{}_I$). Then a reduction of p_1 relative to p_2 made possible by technical change in the production of product 1 is equivalent with a product improvement of 1, such that θ^1 is changed to $\bar{\theta}^1 = \lambda\theta^1, \lambda > 1$, while prices are unchanged.

The proof of the above statement is straightforward. If p_1 is reduced to $\bar{p}_1$, then $z^1(\bar{p})^I > z^1(p)^g$, where $\bar{p}_j = p_j$ except for j = 1. If θ^1 is changed to $\bar{\theta}^1$, $z^1(\bar{p})^g > z^1(p)^g$ where $z^1(\bar{p})^g = \lambda\theta^1 m^g/p_1 = \theta^1 m^g/\bar{p}_1$ if $\lambda/p_1 = 1/\bar{p}_1$. This completes the proof.

Hence, if a process improvement allowing for a price reduction requires the same efforts as a product quality increase, then the two forms of technological change are equivalent.

We may observe that if the price of one commodity, 1, is reduced sufficiently, product 1 with $\theta^1 > 0$ will push another product 2 out from the market by reaching a point $p^*{}_1$ such that

$\theta^1/p^*_1 > \theta^2/p_2$.

This brings us to a generalized case of product superiority.

Remark 8. If product l's vector θ^1 is a linear combination of $\theta^2,...,\theta^k$, then our earlier assumptions leading to (4.1) imply that there is a price p^*_1 and a small scalar $\epsilon > 0$ such that $p^*_1 + \epsilon$ makes $\mu^g_1(...p^*_1 + \epsilon) = 0$ and such that $p^*_1 - \epsilon$ brings about $\mu^g_1(...p^*_1 - \epsilon) = 0$ for all $i = 2,...,k$, and $\mu^g_1(...p^*_1 - \epsilon) > 0$.

Observe first that the μ_i-functions in (4.1) are solutions to (2.11). Remark 8 tells us that with linear dependence the μ_i-functions have discontinuity points which should generate discontinuities of the Y_i-functions describing commodity demand. The statement follows from the linearity of the B operator and the nature of preference functions u^g which imply that at p^*_1

$$\theta^1/p^*_1 = \sum_{i=2}^{k} \mu^g_i(p_2,...,p_k)\theta^i/p_i$$

and the equality sign changes to ">" for $p^*_1 - \epsilon$ and to "<" for $p^*_1 + \epsilon$.

Remark 8 is important indeed if we recognize that product 1 may represent a new or improved product entering the market of an industry. Then it remains to ask the question: should we expect new (and superior) products to capture all buyers in a customer group in one shot? Observations indicate that we should expect a smooth logistic introduction curve. With our current setting this may be obtained by the following assumption. When a superior product 1 attracts a fraction of a customer group 1, we assume that the information about the new product's attributes is spread in an ordinary diffusion process. Let m^g be the budget of the group of active buyers at a point in time, and let M^g be the potential budget of all customers in group g. Then we obtain

$$\dot{m}^g = m^g(1-m^g/M^g) \qquad (4.5)$$

and $p_1 x_1$ will develop accordingly as will μ_1 as a fraction of the total market.

4.3 Product Substitution and Profit Distributions

The superiority of a new product, as described in preceding sections, makes it possible for the product to gradually squeeze out other products at a price that guarantees high profits. At the same time, earlier product vintages will be sold at gradually reduced prices as long as they have not left the market. In particular, the possibility of repetitively introducing new and adjusted products means that the process of improving production techniques may be repeated over and over again, and this will prolong the periods during which we can observe the profit distribution of logistic shape in each industry and industrial subsector as is illustrated in Figure 2.1.

We shall describe this relation between product substitution and profit distributions diagramatically. In figure 4.1 we describe the growing share of a new product $\sigma_1 = x_1/\Sigma_i x_i$. Between time t_a and t_b this share has increased from $\sigma_1(t_a)$ to $\sigma_1(t_b)$.

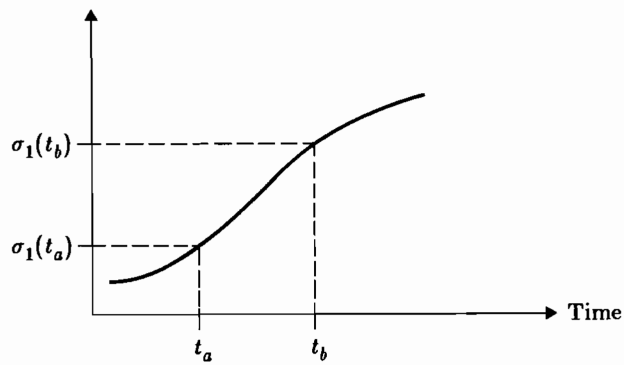


Figure 4.1 Introduction of a new product

In Figure 4.2a and 4.2b we illustrate the effects on the profit distribution of the development described in Figure 4.1. We should also observe that during the gradual increase of σ_1, the production technique will also be improved successively in new capacities introduced to produce product 1.

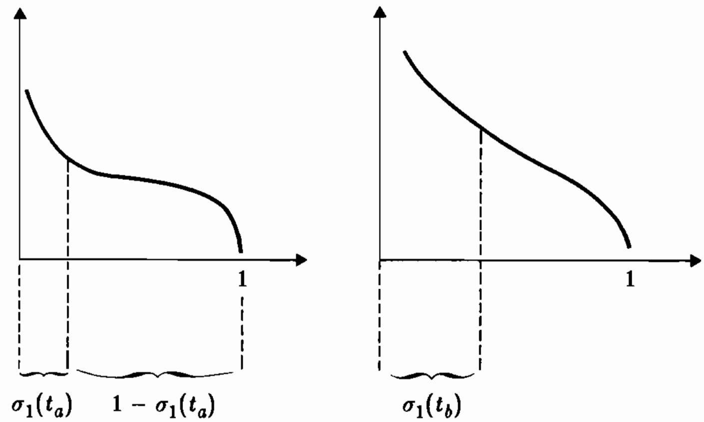


Figure 4.2 Effects on the profit distribution of product substitution

We may now finally comment on the reason for specifying different customer groups. By recognizing that different groups use a product for different functions, it becomes natural to assume that they evaluate the attributes of a product differently. Hence, they may have distinctly different preferences. This helps to explain why "superior" products usually do not conquer the entire market; they are only superior with regard to certain customer groups but not to others.

5. CONCLUSION

In section 4 we noticed the existence of discontinuities which could be the source of dramatic changes in the industrial structure. Such phenomena need thorough investigations. However, one also needs to inquire why we are able to observe clear regularities and invariances in the process of industrial change.

The above remarks are also relevant for the exercises in section 3. It is quite clear that the self-regulating price adjustments in (3.13) and Remark 5, together with (3.5) and (3.6) will produce fluctuations or cycles with cyclic variation in p, x^*_N and E(p). In particular, the combination of the technology-induced price path $\dot{p} = \dot{q} - \lambda \Phi \exp\{-\lambda t\} \bar{c}$ and the excess demand-induced price mechanism in (3.13) together with Remark 5 constitute competing price mechanisms; any interaction between the two will produce oscillations along the time axis. Such phenomena need to be examined with a stronger apparatus than the one used in this outline.

REFERENCES

Berge, C., Espace Topologiques, Dunod, Paris, 1959.

Batten, D. and Johansson B., Industrial Dynamics of the Building Sector: Product Cycles, Substitution and Trade Specialization, in Snickars, F., Johansson, B., Lakshmanan, T.R., (eds.), Economic Faces of the Building Sector, D20:1985, Swedish Council for Building Research, Stockholm, 1985.

Johansson, B., Contributions to Sequential Analysis of Oligopolistic Competition, Memorandum 73, Dept. of Economics, University of Gothenburg, 1978.

Johansson, B, and Strömqvist, U., Rigidities in the Process of Structural Economic Change, Regional Science and Urban Economics, 11; pp. 336-375, 1981.

Johansson, B., and Holmberg, I., A Regional Study of the Distribution of Vintages and Profits of Industrial Establishments: A Stochastic Transition Model, in Albegov, M., Andersson, Å., Snickars, F., (eds.), Regional Development Modelling: Theory and Practice, North-Holland, Amsterdam, pp. 215-227, 1982.

Johansson, B. and Marksjö, B., Interactive System for Regional Analysis of Industrial Sectors, in Nijkamp, P., Rietveld, R., (eds.), Information Systems for Integrated Regional Planning, North-Holland, Amsterdam, pp. 231-249, 1984.

Johansen, L., Production Functions, North-Holland Publishing Co., Amsterdam, 1982.

Lancaster, K., Consumer Demand: A New Approach, Colombia University Press, New York, 1971.

Salter, W.E.G., Productivity and Technical Change, Cambridge University Press, London, 1960.

Schumpeter, J., The Theory of Economic Development, Harvard University Press, Cambridge, Mass., 1934.

Strömqvist, U., Profit Structure and Investment Patterns (in Swedish), SIND 1983:1, Statens Industriverk Liber, Stockholm, 1983.

Chapter 8

Log-Linear Relative Dynamics: Unification of Theories

M. SONIS

1. INTRODUCTION

During the last decade there has been an increasing tendency towards the theoretical unification of different branches of the social sciences. One field of interest in this universal process is the unification of four different theories - innovation diffusion theory, economic utility theory, urban/regional relative dynamics and ecological competition theory - which had previously developed independently without much interaction. This paper aims to review briefly the progress made to date, and to demonstrate the mutual interdependence of the above-mentioned theories. We shall also describe the nature of the ongoing research in this area.

Innovation diffusion theory, which originates from cultural anthropology (Steward 1963), obtained its essential push from Hägerstrand's pioneering work on the modelling of diffusion processes (Hägerstrand 1952). Hägerstrand constructed a theoretical framework based on the complementarity of the contagion and hierarchical effects, and on the important notion of a "mean information field" for adopters of an innovation. But the analytical description of the diffusion process - based on the Verhulst (1838) logistic differential equation - has been very poor until now, something scarcely realized by investigators (see Brown 1981). Important progress came with the consideration of the dynamics among competitive (i.e., exchangeable and mutually exclusive) innovations in time and space (Sonis 1981, 1983a, 1983b).

The derivation of the system of log-linear differential equations of the innovation diffusion process permitted a deeper understanding of the mechanism of an innovation spread: the description of different types of competition between innovations; the study of the asymptotical behaviour of solutions which resulted in the derivation of the competitive exclusion principle; and the introduction of the influence of an active environment, which smooths out the extreme action of the competitive exclusion of innovations, leading to a more balanced distribution of innovations between adopters.

The innovation diffusion process overshadows the process of the individual's choice, which is the main determinant of the speed of innovation spread. While the innovation diffusion theory concentrates on the behaviour of alternative innovations (macro viewpoint), the models of individual choice are connected with the description and explanation of the behaviour of individuals (micro viewpoint). The interconnections between the choice behaviour of individuals and the competitive behaviour of alternative innovations ought to present dual, mutually complementary sides of human activities. Analytically, this duality lies in the fact that the system of log-linear differential equations of the diffusion process has solutions which may be constructed as the choice probabilities for dynamic individual choice models (Sonis 1984a).

Futhermore, the dynamic extensions of the well-known multinomial Logit and Dogit models correspond to a totally antagonistic competition between alternatives, i.e. to the simplest case of competition. The analytical similarity between the static Logit and Dogit models (Domencich and McFadden 1975, Gaudry and Dagenais 1979) and their dynamic counterparts (Sonis 1983c) allows for the use of the same type of statistical evaluation methods in the static and dynamic cases. Conceptually, however, the models are different. The random utility choice Logit and Dogit models are based on the principle of individual utility maximization, while their dynamic counterparts are based on a more realistic accounting of an individual's expectations of future gains, on the incorporation of interactions, imitation and learning processes among adopters, and on the intervention of external forces of an active environment which changes the accessibility of alternatives for the adopters.

In the field of urban/regional dynamics, Hudson (1970, 1972) appears to have been the first to use interaction models for studying population growth and migration in a two-region urban system. Curry (1981) examined the structure of labor occupations under the effects of geographical competition for recruits with the formal use of the ecological competition theory. Recently Dendrinos outlined a new urban/regional dynamics based on a detailed elaboration of interconnections between theoretical urban dynamics, ecological competition theory and bifurcation theory (in a series of papers, resulting in the book "Urban Evolution" by Dendrinos and Mullally 1985).

The interpretation of innovation diffusion theory in terms of interregional competition describes the behaviour of a homogeneous "regional resource" (which in reality can be population, income, production, taxes etc...) relatively distributed and redistributed among several regions. In this scheme interaction among regions describes the input of each region into the relative growth of the others, whereas the competitive exclusion of regions explains the relative regional growth and decline (Sonis 1984b). The competitive exclusion of multiregional dynamics is balanced by the action of external interventions, such as the introduction of national policies or regional development programmes, governmental restrictions on regional autonomy, fiscal and institutional decentralization etc.... The presence of external interventions in regional dynamics generates environmental niches which support and preserve the special type of regional behaviour within an interregional system.

The most interesting fact is that innovation diffusion theory is analytically similar to Volterra's treatment of multi-species "conservative" ecological associations (Scudo and Ziegler 1978). Volterra constructed the conservative ecological dynamics in a manner similar to classical mechanics: the absolute multi-species population growth is caused by a non-zero "self-growth" of each species and by the multi-species interaction, in such a way that the value of total interaction is equal to zero. The main mathematical property underlying Volterra's conservative ecological dynamics is that such dynamics can be derived from Hamilton's variational principle; i.e., that ecological evolution corresponds in fact to a constant value of a quantity Volterra called "vital action". Thus, Volterra's conservative ecological association can be handled with much the same canonical treatment as the classical mechanics of material systems. But biologists soon realized that Volterra's absolute growth conservative ecological dynamics, while being of great mathematical interest, had severe limitations as a representation of nature: for an equilibrium to exist the ecological association must contain an even number of different species (which is difficult to accept from an ecological viewpoint).

Consideration of relative rather than absolute growth, a problem Volterra never addressed, immediately removes the peculiarities in the behaviour of the multi-species ecological association. The relative growth (such as in the innovation diffusion theory or in relative urban/regional dynamics) means a zero self-growth of each species and also the automatic fulfilment of Volterra's conservation condition for the ecological multi-species association. This in turn implies the derivation of relative ecological dynamics from the variational principle of the stationarity of cumulative entropy similar to Volterra's "action" in absolute growth conservative ecological associations (Dendrinos and Sonis 1986). Thus, Volterra's fifty year old studies lend new insights into the construction of a new theory of relative dynamics.

The following is a summary of this paper. In section 2, a review of the mathematical theory of innovation diffusion process is presented. Special attention is given to the substantial interpretation of the analytical properties of the mathematical model. This also has the advantage of demonstrating how previously unnoticed properties of familiar models can be revealed when new methods are applied to them. In section 3, new dynamic individual choice models are derived, as dynamic extensions and generalizations of the static multinomial Logit and Dogit models. An important aspect is the exchange of the principle of utility maximization for a more realistic assumption of interaction among individuals. In section 4, an application of the above-mentioned theories to interregional dynamics is demonstrated. Interregional competition is described as a result of a collision between the action of the entropy maximization principle and the entropy minimization principle for the relative distribution of a "regional resource" among regions; the former presents the competitive exclusion of regions due to competition, while the latter presents the process of regional equalization due to the action of external interventions. Finally, in section 5, the connections between Volterra's "conservative" ecological associations and innovation diffusion dynamics are revealed. As a result, the derivation of the differential equations of relative ecological dynamics is obtained, based on the variational principle of the stationarity of cumulative entropy.

2. INNOVATION SPREAD

To start with, consider a set of N exchangeable, mutually exclusive innovations whose dynamics are governed by a system of log-linear differential equations

$$\frac{dy_i(t)}{dt} = y_i(t) \sum_{j=1}^{N} a_{ij} y_j(t), \qquad i = 1,2...,N, \tag{1}$$

$$\sum_{j=1}^{N} y_j(t) = 1. \tag{2}$$

Here $Y = (y_1, y_2, \ldots, y_N)$ is the vector of relative proportions of adopters of each type from the set of N innovations, and the interaction matrix $A = (a_{ij})$ describes the influence of each innovation on the relative change of all other innovations. The derivation of the system (1), (2) is based on the analysis of information flows between adopters of different innovations, which generate social interaction between individual behaviour of adopters (Sonis 1981). Due to condition (2) of relative growth, the interaction matrix $A = (a_{ij})$ must be antisymmetric:

$$a_{ij} = -a_{ij}\ ; \quad a_{ii} = 0.$$

One can interpret this antisymmetry in the following way; each pair of innovations i and j participates in an antagonistic zero-sum game, with the interaction coefficients a_{ij} being the payoff (expectation of gain) for the ith innovation. This interpretation immediately implies the competitive exclusion principle. If the kth innovation is a "winner" in all antagonistic games against all other innovations such that

$$a_{kj} > 0 \quad \text{for all } k \neq j, \tag{3}$$

then in the long run all adopters will accept only the kth innovation. Thus, in the long run, competitive exclusion will result in the final distribution of adopters of the form

$$y_k = 1; \quad y_j = 0, \ j \neq k,$$

which corresponds to the competitive exclusion vector $E_k = (0,\ldots,0,1,0,\ldots,0)$.

These heuristic considerations are justified by a local stability analysis of the dynamic equilibrium. It is possible to prove that the antisymmetry of the interaction matrix A implies the existence of competitive exclusion equilibrium states E_k, $k=1,2,...,N$ and that condition (3) - the strong positivity of the kth row of the matrix A - is the condition for asymptotical stability of a competitive exclusion equilibrium (Sonis 1983a).

Although this asymptotical behaviour is quite simple, the behaviour of a solution far from the equilibrium is considerably complicated, as it depends on the structure of the interaction among innovations. The simplest type of interaction among innovations is given by the interaction potentials a_k, $k=1,2,...N$, so that the interaction between each pair of innovations i and j is equal to the difference between their potentials:

$$a_{ij} = a_i - a_j.$$

The existence of interaction potentials implies that the total interaction for each closed chain of innovations i,j,k,...,m,i is equal to zero:

$$a_{ij} + a_{jk} + \dots + a_{mi} = 0.$$

This means that each subset m of N innovations participates in a non-cooperative antagonistic zero-sum game, i.e., competition among innovations is totally antagonistic. Moreover, the existence of the interaction potentials leads to the explicit solution of the system (1),(2):

$$y_i(t) = y_i(0) \exp (a_i t)/\sum_j y_j(0) \exp (a_j t), \qquad i=1,2,...,N. \tag{4}$$

In general, when interaction potentials do not exist, the structure of competition corresponds to a different type of cooperative game; moreover, the complete analysis of competition among innovations is based on a breakdown of general competition into totally antagonistic competitions among smaller subsets of innovations (Sonis 1984a).

It is important to underline that in reality the competitive exclusion of innovations is usually smoothed out by the action of an active social and physical environment. This changes the behaviour of innovations and their adopters implicitly, by filtering and directing or intensifying the information flows between adopters and innovations; and explicitly, by physical, social, cultural restrictions and prohibitions, or support and stimulation. The active environment presents the external forces which change the accessibility of the innovations and so changes the distribution of innovations among adopters. The action of such external perturbations can be presented with the help of a stochastic matrix $S = (s_{ij})$, whose coefficients s_{ij} are the probabilities for the adopter of the ith innovation to reject this innovation, and instead adopt the jth innovation under pressure from the active environment.

Therefore the transformation $U(t) = Y(t) S$ converts the competitive exclusion dynamics $Y(t)$ into the more balanced dynamics $U(t)$ influenced by the action of external interventions. Moreover, this transformation simultaneously transfers the competitive exclusion equilibrium states E_k into the balanced equilibrium states $E_k S$ which are the vector-rows of the matrix S.

In the case of the existence of interaction potentials, the totally antagonistic dynamics (4) are transformed into balanced dynamics of the type

$$u_i(t) = \sum_{j=1}^{N} s_{ji} C_j \exp(a_j t) / \sum_{j=1}^{N} C_j \exp(a_j t). \qquad (5)$$

Thus, the influence of external forces by means of the stochastic redistributional matrix opens up the way for the construction of new innovation diffusion models. This is achieved by multiplying the adopter's distribution of some original model by stochastic redistributional matrices representing various types of active environments.

The above-mentioned properties of innovation diffusion dynamics (1),(2) can be transferred to the more general model:

$$\partial y_i / \partial r = y_i \sum_{j} y_j \cdot \partial / \partial r (v_{ij}) \qquad (6)$$

$$\sum_j y_j = 1,$$

where the matrix $V = (v_{ij})$ presents the cumulative influence of the ith innovation on the adoption of the jth innovation. Here the cumulative interaction v_{ij} between innovations i and j depends on changeable characteristics in space and time attributes of these innovations, and various socio-economic characteristics of the adopters; $\partial / \partial r$ denotes the directional derivative in the arbitrary direction r in the space of all explanatory variables and space-time parameters. The transferred properties are: antisymmetry of the cumulative interaction matrix V, the existence of competitive exclusion equilibria E_k, and the antagonistic games interpretation of various types of competition. In the case of the existence of cumulative interaction potentials v_i, such that $v_{ij} = v_i - v_j$, system (6) obtains the explicit solution

$$y_i = y_i^o \exp v_i / \sum_j y_j^o \exp v_j \tag{7}$$

under an initial distribution of adopters $Y^o = (y_1^o, y_2^o, \ldots, y_N^o)$. This solution provides the basis for constructing a dynamic extension of the Logit model of random utility choice (see section 3).

The general (non log-linear) innovation diffusion models

$$\partial / \partial r\, (y_i) = y_i F_{ri}(y_1, y_2, \ldots, y_N), \qquad i=1,2,\ldots,N, \tag{8}$$

$$\sum_j y_j = 1$$

where the non-linear functions F_{ri} can also be transformed into the form (Sonis 1983c):

$$\partial/\partial r\ (y_i) = y_i \sum_j y_j\ \partial/\partial r\ [v_i(y_1,y_2,\dots,y_N) - v_j(y_1,y_2,\dots,y_N)] \tag{9}$$

$$\sum_j y_j = 1.$$

With the help of cumulative interaction potentials $v_i(y_1,y_2,\dots,y_N)$ this form of general innovation diffusion explains many diffusion phenomena and gives a much more realistic description of the innovation spread.

3. DYNAMIC INDIVIDUAL CHOICE

The interconnection between innovation diffusion theory and individual choice behaviour models is based on the interpretation of the distribution of adopters among innovations as an individual's probabilities of alternative choices (innovations). From this viewpoint, the choice decisions are not static but evolve over time, and the individual's choice is essentially influenced by the interaction between individual behaviours. This section sets out to derive the dynamic extensions and generalizations of the well-known static multinomial Logit and Dogit models of random utility choice.

The dynamic extension of the Logit model appears in the case of the innovation diffusion process, and it is associated with the existence of cumulative interaction potentials. In this case, the system of differential equations describing the dynamics of probabilities of individual choice between N alternatives, takes the form

$$\partial y_i/\partial r = y_i \sum_j y_j\ (\partial v_i/\partial r - \partial v_j/\partial r), \qquad i=1,2,\dots,N, \tag{10}$$

$$\sum_j y_j = 1$$

with the explicit solution (7).

By analogy with the static multinomial Logit model (Domencich and McFadden 1975), one may interpret the cumulative interaction potentials v_i as the "systematic" components of individual utilities. Furthermore, the interactions $v_{ij} = v_i - v_j$ are the utilities of transition from the alternative i to j, and $\partial v_i/\partial r$ are the dynamic marginal utilities, which represent the expectations of future gain.

Expression (10) means that the probability y_i increases if

$$\sum_j y_j (\partial v_i/\partial r - \partial v_j/\partial r) > 0.$$

This means that the individual compares the alternative i with all other alternatives j, not by comparing the utilities v_i and v_j only, but also by comparing the dynamic marginal utilities $\partial v_i/\partial r$ and $\partial v_j/\partial r$. Moreover, the consideration of expected transitional utilities only is not sufficient. The individual observes the choice of other individuals and takes into account how many individuals are using the other alternatives. Thus, the term $y_j(\partial v_i/\partial r - \partial v_j/\partial r)$ gives the measure of transitional expected growth in utility, and the degree of influence of adopters of alternative j on the decision to change from alternative i to j.

Such a description of individual choice behaviour necessitates exchanging the principle of individual utility maximization (which is the logical basis for the derivation of the static Logit model), for a more appropriate choice principle which optimizes the sum of dynamic marginal utilities weighted by the distribution of the alternatives among individuals. Thus, an individual cannot be considered as an egoistic "homo economicus" choosing a suitable alternative on the basis of a principle of utility maximization, but rather as a partially altruistic creature, whose imitative and learning behaviour corresponds to the influence and interaction of other individuals. (It is interesting to note that the incorporation into the static Logit model of time, social interaction and utility maximization, transforms it to the mathematically intractable dynamic model (see De Palma and Lefevre, 1983, p. 108)). This conceptualization of individual choice behaviour is also found in a wider frame of general choice models of the type (6), (8) and (9) (Sonis 1983c).

As pointed out in the Introduction, external interventions (the active social and physical environment) alter individual choice behaviour by changing the accessibility of the alternatives. The stochastic redistributional matrix $S = (s_{ij})$, describing the intervention of external forces upon individual choice behaviour, allows for the following interpretation: s_{ij} is the probability of an individual rejecting alternative i and instead adopting alternative j under the influence of external forces.

Introduction of external interventions into the dynamic individual choice models of the types (6), (8), (9), (10) through the transformation U = YS generates new dynamic extensions of the individual choice models, co.responding to different types of external interventions.

For example, we obtain a redistributional stochastic matrix S of the special form

$$S = \begin{bmatrix} (1+s_1)/s & s_2/s & \cdots & s_N/s \\ s_1/s & (1+s_2)/s & \cdots & s_N/s \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ s_1/s & s_2/s & \cdots & (1+s_N)/s \end{bmatrix}$$

where $s = \sum_j s_j$, transforms the dynamic Logit model (10) - or (7) - into the dynamic extension of the static Dogit model (Gaudry and Dagenais 1979):

$$u_i = C_i \exp(v_i) + s_i \sum_j C_j \exp(v_j) / (1+\sum_j s_j) \sum_j C_j \exp(v_j). \tag{11}$$

We can now use an interesting interpretation of this choice behaviour proposed by Ben Akiva for the static Digit model (Ben Akiva 1977). The choice probabilities (11) can be presented as

$$u_j = s_i/(1+\sum_j s_j) + (1/1+\sum_j s_j) C_i \exp(v_i) / \sum_j C_j \exp(v_j);$$

here the first component $(s_i/1 + \sum_j s_j)$ is that part of the choice probabilities u_i which represents the "captivity" of decision-makers with a choice of the alternative i, while the second component of u_i represents the "discretionary" choice of the alternative from all other alternatives.

It is important to stress that the transition from one choice model to another with the help of non-stable stochastic matrices is very useful operationally, because it opens the path for constructing new choice models.

4. MULTIREGIONAL GROWTH AND DECLINE

In this section the main results of mathematical innovation theory are briefly reformulated in terms of interregional competition (Sonis 1984b). The reformulation is restricted because it presents the dynamics of only one homogeneous "regional resource" or "relative regional population"' distributed among N regions. In reality this "relative resource" can be population, labor, capital, income, total national product, governmental expenditures, taxes etc...; and interregional competition is the result of interaction among different types of "relative populations". The construction of the multiregional multiple resources' relative dynamics is the task for future studies.

Assume that some relative population is distributed and redistributed among N regions according to the log-linear system of differential equations

$$dy_i/dt = y_i \sum_j a_{ij} y_j \tag{A1}$$

$$\sum_j y_j = 1, \tag{A2}$$

describing the interregional dynamic redistribution of the proportions $Y = (y_1, y_2, \ldots, y_N)$ of a relative population in each region. The antisymmetric interaction matrix $A = (a_{ij})$ describes the "input" of each region into the relative growth of all other regions, and the coefficients a_{ij} can be interpreted as a regional reward or as a regional burden on the ith region due to the interaction with the jth region. If the ith region is in a preferable position with respect to all other regions,

$$a_{ij} > 0 \text{ for all } j \neq i,$$

then competitive exclusion of regions occurs; i.e., in the long run all relative population will concentrate within the ith region. Thus, spatial competitive exclusion explains the regional disparities and regional growth and decline as a result of interregional competition. In the case where negative coefficients exist in each row of the interaction matrix A, then regional cycles appear.

By looking at the empirical evidence (see, for example, Dendrinos and Mullally 1985) one may find that the equilibrium patterns of multiregional growth do not necessarily confirm the presence of competitive exclusion or regional cycles: balanced final distributions of relative population do exist. This evidence requires the introduction of unbalanced regional external interventions into the models of interregional competition. As possible candidates of external interventions into multiregional competition, one may consider various governmental restrictions on regional autonomy: the introduction of national policies and regional development programmes; conditional grants and interregional compensation funds; institutional fiscal centralization or decentralization etc... The action of external forces can be presented with the help of stochastic matrices $S = (s_{ij})$, describing the processes of the redistribution of relative population among regions due to this action. These redistribution processes act independently from interregional competitive interaction; their mixing and superposition with regional peculiarities generate regional niches which support and preserve the special type of behaviour of regions.

The entropy of the spatial distribution of relative population

$$h = -\sum_j y_j \ln y_j$$

can be used as an indicator of regional disparities. Spatial competitive exclusion is equivalent to <u>entropy minimization</u>, while the process of regional equalization, corresponding to the principles of social justice, leads to <u>entropy maximization</u>; one can claim that the collision between the minimization and maximization of entropy reflects the nature of human progress. The surprising and important feature of cumulative temporal entropy H in regional dynamics, where

$$H = \int_0^T h \, dt$$

is considered in section 5.

One could expect that the above-mentioned properties of multiregional competitive dynamics which apply to one type of relative population could be transferred to the case of multiple population multiregional dynamics. The following model is proposed (Sonis 1985) for the description of the multiregional dynamics of two types of relative resources - the relative distribution of labor $Y_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})$ and the relative distribution of capital $Z_t = (z_{1t}, z_{2t}, \ldots, z_{Nt})$ between N regions - in discrete time:

$$y_{i,t+1} = U_i(Y_t, Z_t) z_{it} \Big/ \sum_{j=1}^{N} U_j(Y_t, Z_t) z_{jt}$$

$$z_{i,t+1} = V_i(Y_t, Z_t) y_{it} \Big/ \sum_{j=1}^{N} V_j(Y_t, Z_t) y_{jt} \qquad (12)$$

$i = 1, 2, \ldots, N,$

where functions U_i, V_i are strictly positive utility functions which represent the cumulative per capita regional utilities, corresponding to the socio-economic characteristics of the population of each region, and the attributes of the regions themselves.

Model (12) is a generalization of the dynamic extension of the multinomial Logit model (7) for the case of two different relative resources. Equations (12) present a new model of relative growth of two different interacting resources, where the spatial regional competitive exclusion principle holds. The equilibrium states (Y^*, Z^*)

for this model are the states $Y^* = Z^* = E_k$ where $E_k = (0,...,0,1,0,...,0)$, k=1,2,...N. This demonstrates the important fact that the action of demographic and economic forces imply total concentration of the labor and capital within only one (kth) region. It must be noted here that the results and the interpretations of sections 2 and 3 also hold for these relative dynamics.

5. RELATIVE ECOLOGICAL DYNAMICS

The mathematical basis for the description of evolution of the population of N species, competing for the same resource, was constructed by Volterra in the early decades of this century. Volterra considered the absolute population growth of the multispecies' ecological association to be influenced by the multiple species interaction in the form (Scudo and Ziegler 1978):

$$dx_i/dt = x_i(a_i + \sum_{j=1}^{N} c_{ij}x_j/b_i), \qquad i = 1,2,...,N, \tag{13}$$

where $x_1, x_2, ..., x_N$ are the amounts of each type of species, a_i is the rate of "self-growth", b_i is an "average weight" of species and c_{ij} are the interaction coefficients.

The key notion of Volterra's elaboration is the notion of a "conservative ecological association" for which

$$\sum_{i=1} \sum_{j=1} c_{ij}x_jx_i = 0. \tag{14}$$

There is no need for this restriction of antisymmetry to be imposed on relative ecological associations possessing total zero growth:

$$\sum_i x_i = P = \text{const.} > 0. \tag{15}$$

Indeed, let us consider the relative ecological dynamics with total zero growth (12),(15). (Volterra avoided the consideration of this particular case of ecological dynamics.) The introduction of new variables $y_i = x_i/P$ and new interaction coefficients $a_{ij} = a_i + Pc_{ij}/b_i$ result in the innovation diffusion dynamics (1),(2). Moreover,condition (2) for relative growth is equivalent to Volterra's conservation condition

$$\sum_i \sum_j a_{ij} y_i y_j = 0$$

because

$$\sum_i \sum_j a_{ij} y_i y_j = \sum_i y_i(\sum_j a_{ij} y_j) = \sum_i dy_i/dt = d/dt(\sum_j y_j) = 0.$$

Therefore, it is possible to interpret the relative ecological dynamics as a special case of Volterra's conservative ecological association, with a zero self-growth ($a_i = 0$) and a zero aggregative (multiregional) growth (P = const.). The immediate consequence of such an interpretation is the possibility to employ Volterra's analytical methods and results in relative ecological dynamics (and at the same time avoid the peculiarities of absolute growth conservative ecological associations).

Only one important fact related to this will be mentioned here: the derivation of the differential equations (1),(2) of relative dynamics from the variational principle of the Hamilton type (Dendrinos and Sonis 1986).

Let us consider the cumulative proportions of relative population

$$Y_i(t) = \int_0^t y_i(t)\, dt$$

(the analogue of Volterra's "quantities of life"), and the integral of cumulative action

$$\int_0^T (-2 \sum_i y_i \ln y_i + \sum_i \sum_j a_{ij} y_i y_j)\, dt . \qquad (16)$$

The analogue of Hamilton's principle of a stationary cumulative action means that the first variation of action in (16) vanishes, giving rise to the system of Euler differential equations; for integral (16) the Euler conditions coincide with the system (1),(2) representing the log-linear relative dynamics. The most important fact is that the stationary value of cumulative action (16) turns out to be the cumulative entropy

$$\int_0^T \left(-\sum_j y_j \ln y_i\right) dt$$

for the relative log-linear dynamics over the time horizon T. This is of particular interest to regional analysis, since entropy over space evolves to a minimum in the asymptotically stable equilibrium states (competitive exclusion). However, over time the cumulative action (16), which includes the accumulation of the spatial entropy and the interactions, merges into the stationary value (weak maximum) of cumulative entropy.

6. CONCLUSION

This paper has argued that four different theoretical frameworks, which up to now have been developed in an independent manner (namely innovation diffusion, individual utility theory, urban/regional relative dynamics, and the mathematical theory of ecological competition) have the same conceptual basis: log-linear relative dynamics.

There were two objectives in this paper. The first objective was to demonstrate the mutual interdependence and complementarity of these conceptual frameworks. The second objective was to present the most significant implications of the unification of the different theories to date: the presentation of the innovation spread as competition among antagonistic innovations; the duality between the competitive behaviour of innovations and the individual's choice; the derivation of dynamic counterparts of the multinomial Logit and Dogit choice models and their generalizations; the description of regional growth and decline as a result of multiregional competition; the construction of relative ecological dynamics, a problem ecologists had never addressed; and, finally, the derivation of a system of log-linear equations for relative dynamics on the basis of the variational principle of stationarity in cumulative entropy in relative population distributions.

One may conclude that the unifying view-point of the log-linear relative dynamics presents a new way of looking at old problems and leads to essentially new results.

ACKNOWLEDGEMENT

I thank Dimitrios Dendrinos for reading, correcting and commenting on a draft of this paper, and especially for presenting it on my behalf at the 5th International Conference on Mathematical Modeling at University of California, Berkeley, July 1985.

REFERENCES

Ben Akiva, M.E., "Choice models with simple choice set generating processes ", Working paper, Centre of Transportation Studies, Massachusetts Institute of Technology, 1977,

Brown, L.A., Innovation diffusion - a new perspective, Methuen, London, 1981.

Curry, L., "Division of labor from geographical competition", Annals of the Association of American Geographers, 71, 1981, pp. 133-165.

Dendrinos, D.S. and Mullally, H., Urban evolution: Studies in the mathematical ecology of cities, Oxford University Press, New York, 1985.

Dendrinos, D.S. and Sonis, M., "Variational principles and conservation conditions in Volterra's Ecology and in urban relative dynamics", Journal of Regional Science, 26, 1986, pp. 359-377.

De Palma, A. and Lefevre, C., "Individual decision-making in dynamic collective systems", Journal of Mathematical Sociology, 9, 1983, pp. 103-124.

Domencich, T. and McFadden, D., Urban travel demand: a behavioral analysis, McGraw-Hill, New York, 1975.

Gaudry, G. and Dagenais, M., "The Dogit model"', Transportation Research, 13B, 1979, pp. 105-111.

Hägerstrand, T., Innovationsfoerloppet ur korologisk synpunkt, Lund, C.W.K. Gleerups, 1952. English translation: A. Pred, Innovation diffusion as a spatial process, University of Chicago Press, Chicago, 1967.

Hudson, J., "Elementary models for population growth and distribution analysis", Demography, 7, 1970, p. 3.

Hudson, J., Geographical diffusion theory. (Studies in Geography, 19), Northwestern University, Evanston, Ill., 1972.

Scudo, F. and Ziegler, J., (eds.), The Golden Age of theoretical ecology: 1923-1940. Lecture Notes in Biomathematics, 22, Springer Verlag, New York, 1978.

Sonis, M., "Diffusion of competitive innovations", in Vogt, W. and Nickle, N., (eds.), Modeling and Simulation, 12, 1981, pp. 1037-1041. Instrumental Society of America, School of Engineering, University of Pittsburgh, Pittsburgh PA.

Sonis, M., "Competition and Environment: a theory of temporal innovation diffusion process", in Griffith, D. and Lea, A., (eds.), Evolving Geographical Systems. Martinus Nijhoff, The Hague, 1983a, pp. 99-129..

Sonis, M., "Spatial-temporal spread of competitive innovations: an ecological approach", Papers of the Regional Science Association, 52, 1983b, pp. 159-174.

Sonis, M., "Dynamic discrete time multinomial Logit and Dogit models", Paper presented at XXIII European Congress of Regional Science Association, August-September 1983c, Poitiers, France.

Sonis, M., "Dynamic choice of alternatives, innovation diffusion, and ecological dynamics of the Volterra-Lotka model", London Papers in Regional Science, 14, 1984a, pp. 29-43.

Sonis, M., "Regional growth, regional decline and decentralization", Paper presented at Regional Science Summer Institute "Problems and possibilities of regional decentralization", August 1984b, Bamberg, Germany.

Sonis, M., "Labor-capital multiregional dynamics". Paper presented at NATO Advanced Studies Institute "Transformations through Space and Time", August 1985, Hanstholm, Denmark

Steward, S., Theory of cultural change, University of Illinois Press, Urbana, 1963.

Verhulst, P., "Notice sur la loi que la Population Suit Dans son Accroissement", Correspondence Mathématique et Physique, 10, 1838, pp. 113-121.

PART B: PERIODIC CYCLES AND STRUCTURAL ADJUSTMENT

Chapter 9

The Schumpeter Clock

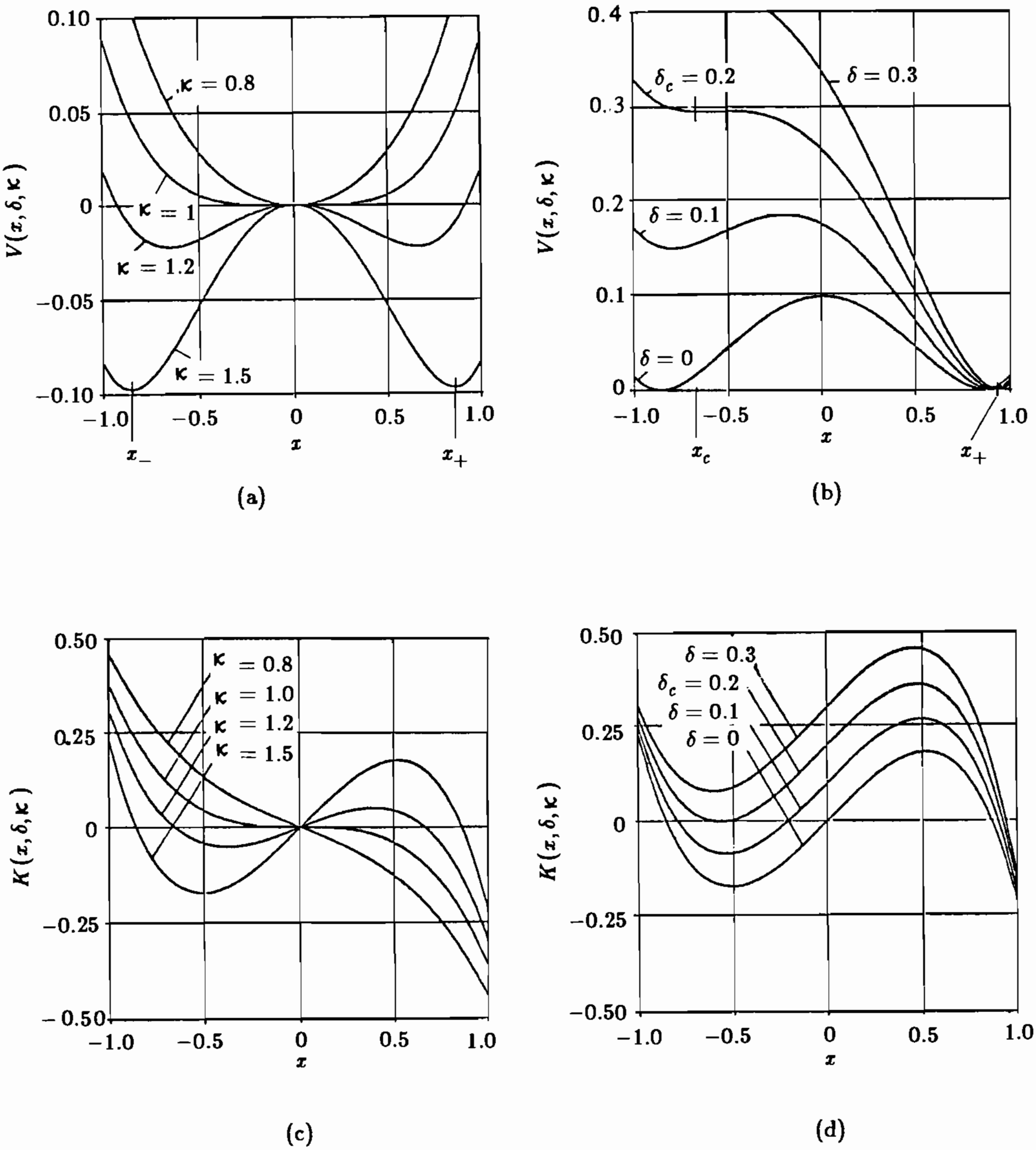
G. HAAG, W. WEIDLICH & G. MENSCH

1. INTRODUCTION

It is a fundamental fact that the non-equilibrium behaviour of an economy consists of long-term and short-term cycles (Keynes 1973; Hicks 1956) and fluctuating components. As a first approximation, the long- and short-term phenomena can be treated separately (Figure 1). In this contribution, we shall focus on short-term motions of the economy which manifest themselves in the macroeconomic variables. We aim to provide a partial theory for the non-equilibrium motion of an industrial system of nations or regions.

The model differs from other existing models of industrial fluctuations in many of its design principles (Goodwin 1950; Chang and Smyth 1950). It is called "the Schumpeter Clock" here according to a proposal from Goodwin (1951), since its moving parts, driving mechanism and control devices are typically Schumpeterian and not, as in the case of other models, typically neo-classical or neo-keynesian. This post-Schumpeterian mathematical theory of short-term cycles is based on microeconomic concepts of the Schumpeterian variety (Schumpeter 1961). It refers to the functioning of the Schumpeter Goods Sector, which is essentially comprised of industry, and those parts of agriculture and services which are operating similarly to industrial organizations. Furthermore, this economic theory refers to dynamic change in major statistical units such as whole nations or regions, and is thus oriented towards the institutional user of applied macroeconomic theory.

The model is primarily designed to take into account hard-driving, microeconomic



<u>Figure 1</u> (a, b) Typical graphs of the configuration potential $V(x,\delta,\kappa)$
(c, d) The associated configuration driving forces $K(x,\delta,\kappa)$

forces and powerful supply side checks and balances, which combine to create short-term oscillations. The soft-driving macroeconomic forces, and the weak demand side checks and balances, are assumed to be mediated through microeconomic behaviour. In particular, the model operationalizes the Schumpeterian notion of the Prime Mover, i.e., innovators and imitators, who create and propagate microeconomic differences. The heterogeneity among products and production processes constitutes the force field for the creative and destructive energies which hold the economic process in motion. (In this sense, Schumpeter's Process of Creative Destruction can also be understood as the Process of Monopolistic Competition. E.H. Chamberlin has shown that this concept of the product as an economic variable leads Toward a More General Theory of Value (1957)).

Differences among products come into play at the disaggregate levels (firms, markets) of the economic system. The creation of such differences - leading to competitive advantages among rival producers - is the objective of the strategic investments of entrepreneurs. We classify strategic investments I(t) according to their respective competition purposes as expansionary E(t) and rationalizing R(t). These investments move the extensive and intensive frontiers of the economy.

The link between the newly introduced macroeconomic investment structure index and the microeconomic investor's configuration index can be made by introducing methods out of the field of synergetics (Weidlich and Haag, 1982). The investment structure index is a measure for the composition of the aggregate strategic investment, namely, of its expansionary (E) and rationalizing (R) shares. The investors' configuration describes all investors' propensities as either expansionary or rationalizing, as they make their strategic investments predominantly either for extensifying or for intensifying purposes; and the investors' configuration index describes the expansionary bias, or rationalization bias, prevalent in the combined population of investors in any one period of time. Because of the tight relationship between the investors' configuration index and the investment structure index, short-term shifts in bias for either expansionary or rationalizing investments translate directly into swings in the composition of total realized investment in the whole economy. E-type investments extend the capacity of firms, industries, and the economy as a whole. R-type investments have ambivalent macroeconomic effects. As they intensify cost factors, rationalizing investment projects have contractionary ramifications for the supply side of the economy on the whole. Thus, micro-shifts in individual investment propensities combine in generating up- and down-swings of the economy.

The main argument for the cyclicity of industrial (short-term) development presented here is derived from the notion of a dynamics of the shifts between differentiation (innovation) and conformative behaviour (imitation). Over the years, the majority of industrial investors repeatedly shift from a predominantly expansionary investment portfolio to a rationalization bias, with innovators and pioneering entrepreneurs in search of monopoly profits taking the lead in the anticyclical redirection of investment strategies. This "band-waggon" effect results in industrial fluctuations.

In the following sections, equations of motion, which quantify the dynamics of economic change, will be set up and solved. The model simultaneously connects the relevant micro-economic and macro-economic concepts, although it is certainly partial, with its focus on the theory of industrial fluctuations, and in its abstraction of the impact of other substantive areas of economic theory.

According to Gold's managerial decision coefficient model, there are only two principal option types; those directed at either increasing sales volume S, or decreasing the cost level C, in return-on-investment (r.o.i.)

$$r.o.i = \frac{S-C}{\text{output}} \cdot \frac{\text{output}}{\text{capacity}} \cdots \frac{\text{capacity}}{\text{fixed investment}} \tag{1}$$

Thus, both expansionary (E-type) and rationalizing (R-type) investment can increase the r.o.i. by increasing the profits (S-C).

Accordingly, the total volume of strategic investment (Kalecki 1935)

$$I(t) = E(t) + R(t) \tag{2}$$

is defined as all fixed capital investment, replacement deducted, which is considered to be tactical, where

$E(t)$ is the <u>volume of expansionary investment</u> at time t and
$R(t)$ is the <u>volume of rationalizing investment</u> at time t. (3)

We propose that there are considerable fluctuations in the volume of expansionary and rationalizing investment as a result of entrepreneurial innovation and imitation activities. The proportional distribution of the shares of $E(t)$ and $R(t)$ in $I(t)$ shifts with time. As these fluctuations take place around long term average paths ($E_0(t)$ and $R_0(t)$) of the expansionary and rationalizing investment, it is first appropriate to decompose $E(t)$ and $(R(t)$

$$\left.\begin{aligned} E(t) &= E_0(t) + B(t) \\ R(t) &= R_0(t) - B(t) \end{aligned}\right\} \quad (4)$$

where the oscillating shift $B(t)$ around the average values E_0 and R_0 is of interest. While the long term behaviour of E_0 and R_0 displays positive semidefinite quantities by definition, $B(t)$ can only vary within the range

$$-E_0 < B(t) < R_0. \quad (5)$$

An investment structure index is now defined as

$$Z(t) = [E(t) - R(t)] / [E(t) + R(t)] = [E(t) - R(t)] / I(t) \quad (6)$$

where $Z(t)$ varies within $-1 < Z(t) < +1$. Inserting (4) into (6), $Z(t)$ decomposes into

$$Z(t) = Z_0(t) + z(t) = (E_0 - R_0) / I(t) + 2\, B(t) / I(t). \quad (7)$$

The performance of the Schumpeter Clock will be demonstrated by observing the non-equilibrium motion of the investment structure index $Z(t)$, or preferably, its fluctuating part $z(t)$, for the industrial sector in the Federal Republic of Germany during the period 1956-1979. This will be then explained by using micro-economically determined and supply side factor reinforced shifts of bias in the overall investment activities.

In principle, an industrial firm may plan and/or undertake any number of projects at a given time t, but since the "investor" is defined as the decision-making unit, one project per firm will be assumed. The total number of investment projects is large, say $2N \gg 1$, for it is convenient - but not decisive - for the equations to have an even number of projects. For simplicity, it will also be assumed that all projects have the same financial volume . This assumption has no effect on the mean value considerations of this chapter; differences in the size of projects would only influence the variances of the model parameters.

A fictitious "neutral" investor who behaves according to the average long term investment trend will be considered first. His investment project of volume $i = I/2N$ is composed of expansionary investment e_0 and rationalizing investment r_0 as follows:

$$i = I/2N = e_0 + r_0 \quad \text{where} \quad e_0 = E_0/2N, \quad r_0 = R_0/2N. \tag{8}$$

Real investors, however, behave differently to the neutral investor: there are E-type investors who favour the expansionary investment type rather than the average trend. Their projects of financial volume i are constructed as follows:

$$i = e_E + r_E \quad \text{where}$$

$$e_E = e_0 + b \quad \text{and} \quad r_E = r_0 - b, \quad \text{with } b > 0. \tag{9}$$

Evidently b is the surplus of expansionary investment in comparison with the neutral case.

On the other hand, R-type investors favour the rationalizing investment type rather than the average trend. The E/R type shares of their projects are such that

$$i = e_R + r_R \quad \text{where}$$

$$e_R = e_0 - b \quad \text{and} \quad r_R = r_0 + b, \quad \text{again with } b > 0. \tag{10}$$

b is now the surplus in rationalizing investment with respect to the neutral case. For simplicity, the same value for b in (9, 10) will be taken.

At a given point in time $n_E(t)$ E-type investors and $n_R(t)$ R-type investors, whose total number is 2N, where:

$$n_E(t) + n_R(t) = 2N(t) \tag{11}$$

are assumed.

The two numbers in the vector $\{E(t), R(t)\}$ characterize the investment structure at time t according to investment volume, and the two numbers $\{n_E(t), n_R(t)\}$ characterize the investors' strategic investment activities at time t (by head count). $\{n_E(t), n_R(t)\}$ is denoted as the <u>investors' configuration</u>.

The integer

$$n(t) = \{n_E(t) - n_R(t)\} / 2 \qquad \text{where} \qquad -N \leqq n(t) \leqq N \tag{12}$$

increases or decreases by one if the investors' configuration changes according to transitions

$$\{n_E, n_R\} \rightarrow \{n_E + 1, n_R - 1\} \tag{13a}$$

or

$$\{n_E, n_R\} \rightarrow \{n_E - 1, n_R + 1\} \ , \tag{13b}$$

i.e. if an R-type investor becomes an E-type investor or vice versa. Multiple unit motions (band-wagon effects) within the investors' configuration are also possible, for instance

$$\begin{aligned} \{n_E, n_R\} &\rightarrow \{n_E + d, n_R - d\} \\ n &\rightarrow n + d \, . \end{aligned} \tag{14a}$$

Such transitions may indicate series of synchronized <u>product</u> variations or innovations, as often observed and described by the heuristic regularity usually called the product life cycle. Similarly,

$$\{n_E, n_R\} \rightarrow \{n_E - d, n_R + d\} \qquad (14b)$$

$$n \rightarrow n - d$$

indicates series of synchronized <u>process</u> improvements or process rationalizations.

Instead of $n(t)$, a normalized variable, the <u>investors' configuration index</u>, can be employed so that

$$x(t) = \{n_E(t) - n_R(t)\} \; / \; \{n_E(t) + n_R(t)\} = n(t) \; / \; N(t) \qquad (15)$$

with

$$-1 \leqq x(t) \leqq 1.$$

The relation between the investors' configuration index (15) and the investment structure index (6), characterizing strategic investment, now follows unambiguously by combining the postulated equations. The total expansionary investment $E(t)$ and the total rationalizing investment $R(t)$ at time t are given by

$$E(t) = n_E(t)\, e_E + n_R(t)\, e_R \qquad (16)$$

$$R(t) = n_E(t)\, r_E + n_R(t)\, r_R\, . \qquad (17)$$

Inserting (8-10) as well as (11, 12, 15) it follows that

$$E(t) = [n_E(t) + n_R(t)]\, e_0 + [n_E(t) - n_R(t)]\, b$$

$$= 2\, Ne_0 + 2\, n(t)b = E_0 + 2\, Nb\, x(t) \qquad (18)$$

and

$$R(t) = [n_E(t) + n_R(t)]\, r_0 - [n_E(t) - n_R(t)]\, b$$

$$= 2\, Nr_0 - 2\, n(t)b = R_0 - 2\, Nb\, x(t) \qquad (19)$$

$$Z(t) = (E_0 - R_0) \,/\, I(t) + [4\, Nb \,/\, I(t)]\; x(t) = Z_0 + z(t) \qquad (20)$$

or

$$z(t) = [4\, Nb \,/\, I(t)]\; x(t) = g\, x(t) \quad \text{with} \quad g(t) = [4\, N(t)b(t)] \,/\, I(t). \qquad (21)$$

The result (21) shows that the fluctuating part $z(t)$ of the investment structure index

Z(t) is proportional to the investors' configuration index x(t). Thus, oscillations of the investors' configuration will show up in oscillations of the investment structure index. The variable g(t) provides the link between the normalized x(t) and the non-normalized z(t).

The discussion of the investors' strategic choice set indicated that the model is consistent with the traditional profit seeking hypothesis. The investigation of the investors' configuration reveals that the model is also consistent with the rational expectation hypothesis. "Rational" expectations are model based, and the life cycle model and progress function model are frequently used by investors to form expectations about the timing of their rivals' next improvements, (Barzel 1968; Kamien and Schwartz 1972) i.e. of the current best practice frontier.

Rosenberg (1976) also pointed out the rationality of speeding-up or delaying product or process innovations under conditions of rivalry. This is done in the expectation of breakthroughs and rapid successions of product or process innovations, which hit the market as a series of capacity expanding or cost reducing investments. Also, in the "awareness context" (Glaser and Strauss 1964) of industrial investors "a firm must incur some positive expense just to maintain a constant level of production cost or efficiency" (Flaherty 1980).

This is especially true in times of cost inflation when the goal is one of "reducing cost in total, not particular costs such as labour costs or capital costs" (Salter 1960). This micro-economic theory of macro-economic contractions also suggests that - in times of cost inflation in which industrial firms try to combat a strong rationalization bias in their overall investment activities - heavy investment does not create higher employment, as the Phillips curve would suggest, but "stagflation".

2. THE EQUATIONS OF MOTIONS FOR THE INVESTORS' CONFIGURATION

It will be assumed that short term industrial investment cycles come about through dynamic interaction between industrial investors, or between two interrelated components of investment behaviour: a) the decision behaviour of managers and entrepreneurs who are making strategic choices about the kind of industrial investments to be implemented, and b) the actual total volume and composition of realized investments. As these components are embodied in the investors' configuration, changes in the numbers $n_E(t)$ and $n_R(t)$ of investors undertaking expansionary or rationalizing investments can be taken as a proxy for the real socio-economic causes of the time path of all industrial activity. Changes in the investors' configuration $\{n_E(t), n_R(t)\}$ result in changes in the rate and direction of industrial investment, as indicated by the investment structure index Z(t). These changes come about as investors change their propensities for E- and R-type investments in view of market opportunities, the revealed preferences of other investors, and their supply side conditions.

Thus, changes in the industrial economy can be formulated as equations of motions of these two components of investment behaviour. The equation of motion for the investors' configuration and the equation of motion for the investors' propensities, as presented in the next two sections, constitute essential parts of the Schumpeter Clock model being presented.

Is has been shown that the transition from one investors' configuration $\{n_E, n_R\}$ to another can be a single unit motion connected with a product innovation (13a) or a process innovation (13b) of one investor. It could also be a multiple unit motion (14a and b) which is most often an imitation process. As innovation is always, and imitation is sometimes, investment under uncertainty, it is doubtful that a deterministic modelling of the motion of the investors' configuration will succeed. But since all investment is risky, the stochastic approach is appropriate. Therefore the well founded master equation formulation will be adopted.

The micro-economic approach to changes in the investors' configuration $\{n_E, n_R\}$ adopted incorporates the notion of individual transition probabilities (per unit of time) for investors turning from an R-type investment to an E-type investment, and vice versa. These individual transition probabilities are denoted by:

$$p_{E \leftarrow R}[n_E, n_R] \equiv p\uparrow(n) = \text{probability per unit time for turning from } \underline{R}\text{-type to } \underline{E}\text{-type investment} \quad (22)$$

$$p_{R \leftarrow E}[n_E, n_R] \equiv p\downarrow(n) = \text{probability per unit time for turning from } \underline{E}\text{-type to } \underline{R}\text{-type investment.} \quad (23)$$

The individual transition probabilities yield the total probabilities for changes of the entire investors' configuration. The transition

$$\{n_E, n_R\} \rightarrow \{n_E+1, n_R-1\}$$

takes place with the total transition probability

$$w\uparrow(n) = n_R \; p\uparrow(n) = (N-n)p\uparrow(n) \tag{24}$$

as there are n_R investors who could turn from R-type to E-type investment with individual probability $p\uparrow(n)$. Analogously, the transition

$$\{n_E, n_R\} \rightarrow \{n_E-1, n_R+1\}$$

takes place with the total transition probability

$$w\downarrow(n) = n_E \; p\downarrow(n) = (N+n) \; p\downarrow(n) \tag{25}$$

as there are n_E investors who could turn from E-type to R-type investment with individual probability $p\downarrow(n)$. These transition probabilities (24, 25) of single unit motions of the investors' configuration $\{n_E, n_R\}$ are the fundamental inputs for formulating the master equation.

The master equation describes the motion of the probability distribution over the investors' configuration. This distribution function is denoted as

$$P[n_E, n_R ;t] \equiv P(n;t) \quad \text{for} \quad -N \leqq n \leqq N. \tag{26}$$

By definition, the value of this function of n and t is the probability that at time t the investors' configuration $\{n_E, n_R\}$ is realized; the configuration relevant variable of the function $P(n;t)$ is $n=(n_E - n_R)/2$. Because one of the configurations $\{n\}$ is always realized, the sum of the probabilities has to be 1:

$$\sum_{n=-N}^{N} P(n;t) = 1. \tag{27}$$

The master equation is derived by a simple probability balance consideration: the probability P(n;t) of the configuration n can <u>increase</u> by means of transitions from either one of the neighbouring configurations (n-1) or (n+1) into configuration n, while P(n;t) may simultaneously <u>decrease</u> by transitions from configuration n into these neighbouring configurations. The transition probabilities per time unit for these processes have already been introduced in (24, 25). Incorporating them into the balance consideration immediately leads to the master equation

$$dP(n;t) / dt = [w\uparrow(n-1)P(n-1;t) + w\downarrow(n+1)P(n+1;t)] - [w\uparrow(n)\,P(n;t) + w\downarrow(n)\,P(n;t)]\,. \tag{28}$$

The master equation (28) represents $2N + 1$ coupled linear difference differential equations for P(n;t) with $n = -N, -N+1, \ldots, N-1, N$, which are not easily solved in the general case when N is large. However, if it is assumed that the P(n;t) are sharply peaked and uni-modal around their mean values $\langle n \rangle_t$, then a set of closed equations of motion for the mean values can be derived. These are defined by

$$\langle n \rangle_t = \sum_{n=-N}^{N} nP(n;t) \tag{29}$$

Taking the time derivative of (29) and inserting the master equation (28) in the rhs, one obtains

$$d\langle n \rangle_t / dt = \sum_{n=-N}^{N} n\,[dP(n;t)] / dt = \sum_{n=-N}^{N} [w\uparrow(n) - w\downarrow(n)]\; P(n;t). \tag{30}$$

The rhs of (30) is the mean value $\langle w\uparrow(n) - w\downarrow(n)\rangle_t$ of $w\uparrow(n) - w\downarrow(n)$. Equation (30) is not a closed equation for the mean value $\langle n \rangle_t$ because the full distribution P(n;t) has to be known in order to calculate $d\langle n\rangle_t/dt$. But under the assumption that $P(n;t)$ remains sharply peaked and uni-modal around its mean value, an approximate closed equation of motion for the mean value $\langle n\rangle_t$ can be obtained as

$$d\langle n\rangle_t / dt = w\uparrow(\langle n\rangle_t) - w\downarrow(\langle n\rangle_t). \tag{31}$$

The mean value equation for x, the investors' configuration index, upon transition to variable $x = n/N$, (see (15) in (31)) then reads:

$$d\langle x\rangle_t / dt = K(\langle x\rangle_t) \tag{32}$$

$$K(\langle x\rangle_t) = 1/N\,[w\uparrow(\langle n\rangle_t) - w\downarrow(\langle n\rangle_t)] =$$
$$(1-\langle x\rangle_t)p\uparrow(N\langle x\rangle_t) - (1+\langle x\rangle_t)p\downarrow(N\langle x\rangle_t). \tag{33}$$

From this point on the clumsy mean value brackets will be omitted and $x(t)$ written for $\langle x\rangle_t$, thus (32) is

$$dx(t)/dt = K(x(t)). \tag{34}$$

According to (33), K depends on the actual configuration $x(t)$ of all the actors (investors) in the system at time t, and the transition probabilities for change of this configuration $p\uparrow(x)$ and $p\downarrow(x)$, which in turn depend upon all investors' propensities, which shall be parametrized by δ and κ. Thus the driving force K, see (33), may be written as a function of x, δ and κ as

$$K(x;\delta,\kappa) = (1-x)p\uparrow(x;\delta,\kappa) - (1+x)p\downarrow(x;\delta,\kappa). \tag{35}$$

A fairly general constructive specification of these transition probabilities which has already been shown to be sufficiently flexible takes the form:

$$\begin{aligned} p\uparrow(x;\delta,\kappa) &= \nu\,\exp(\delta+\kappa x) \\ p\downarrow(x;\delta,\kappa) &= \nu\,\exp[-(\delta+\kappa x)] \end{aligned} \tag{36}$$

(Weidlich and Haag, 1983).

The technical parameter ν is a frequency which relates, according to (49), the abstract dimensionless time τ in formal modelling to real time t in empirical work, when the model is being applied to real economic data. The trend parameter δ is referred to as The Alternator. It initiates the reversal of strategic bias. To see this, first assume an increasingly positive value of δ in (36). It increases the value of $p\uparrow$, whereas the value of $p\downarrow$ is decreased. Therefore, a positive alternator favours individual strategy change over to E-type investment, and it disfavours transitions to R-type investment. The inverse holds for negative δ. In our model the alternator is viewed as a function of time $\delta(t)$. A dynamic equation of motion for the alternator shall be set up in section 3. It alternates between positive and negative values over time, and its oscillations produce periodic shifts of the probabilistic decision behaviour of investors. In other words, the alternator governs the propensity to switch the strategic direction of innovation.

The trend parameter κ represents the propensity to imitate. Thus, it is referred to as The Coordinator. It amplifies the strategic shifts. To see this, let us first focus upon a situation with $n_E > n_R$ or $x > 0$. This means that the majority of firms are currently investing with expansionary bias. Then, according to (36), the transition probability $p\uparrow$ of an R-type investor to switch over to E-type investment is larger than the transition probability $p\downarrow$ of an E-type investor to change from E-type to R-type investment. The converse is true in a situation where $n_E < n_R$, or $x < 0$. In both cases, however, the transition probability for change to the investment type of the majority is larger than that for a change to the minority type. Hence, the incumbent majority is stabilized and even extended by the effect of κ. This trend increases for increasingly positive κ. In other words, the coordinator represents the investors' inclination to conform to the behaviour of the majority of firms at the time. The coordination effect manifests itself as a synchronization effect of investments of the same type undertaken by a majority of investors. In general, the effects of the alternator and the coordinator superpose and thus determine the transition probabilities at any point in time. Since δ and x depend on time, the transition probabilities of investors are themselves functions of time. They simultaneously cause strategic changes, and depend on the effects of the changes which manifest themselves in the economic data.

Inserting (36) in (35) and using the standard definitions of the hyperbolic sine and cosine, a specific expression for the driving force

$$K(x;\delta,\kappa) = 2\nu[\sinh(\delta+\kappa x) - x\cosh(\delta+\kappa x)] \qquad (37)$$

is obtained, which in turn yields an explicit form for the mean value equation of motion of the investors' configuration index

$$dx(t) / dt = K(x;\delta,\kappa) \qquad (38)$$

by using (37) on the rhs of (38).

The dynamic behaviour of x can also be described in terms of a "potential" which stands in close relation to the driving force K of x such that

$$K(x;\delta,\kappa) = -\partial V(x;\delta,\kappa) / \partial x, \qquad (39)$$

which may be regarded as being the pace maker of the Schumpeter Clock. Integrating the force (37) yields the "configuration development potential"

$$V(x;\delta,\kappa) = (2\nu/\kappa^2)\,[\kappa x\sinh(\delta+\kappa x) - (1+\kappa)\cosh(\delta+\kappa x)]. \qquad (40)$$

In the upper part of Figure 1, numerical values for $V(x;\delta,\kappa)$ are depicted:

(a) for $\delta = 0$ and different values of κ
(b) for $\kappa = 1{,}5$ and different of values δ,

whereas in the lower part of Figures 1c and d, the quantified driving forces $K(x;\delta,\kappa)$ associated with the cases a and b, respectively, are shown. Comparing Figure 1a with c, and b with d for the same sets of parameters κ, δ it is seen that the force $K(x;\delta,\kappa)$ drives x into the minima of the potential $V(x;\delta,\kappa)$. In case a with $\kappa = 1.5$ for instance, the motion ends at the positive minimum $x = x_+$ (expansionary bias of the investors' configuration index x) or at the negative minimum at $x = x_-$ (rationalization bias of x). In case b with $\delta = 0.3 > \delta_c$, the potential has one minimum only at $x = x_+$

and the motion ends with the strong expansionary bias $x = x_+$.

It may be supposed that the motion of the investors' configuration index x always ends at a stable (expansionary or rationalizing) bias and that no Schumpeter Clock begins to tick. But this is only true as long as δ and κ are kept constant with time. In section 3, however, an equation of motion for the alternator $\delta(t)$ will be set up. In becoming a dynamic variable, the alternator $\delta(t)$ shifts and transforms the whole shape of the configuration development potential $V(x; \delta,\kappa)$ and of the driving force $K(x; \delta,\kappa)$. Hence x may never come to rest; in fact a periodicity in the motion of $x(t)$ can result from the coupling of the motion of $\delta(t)$ with that of $x(t)$, and so oscillatory shifts in the variable x in the course of this coupled motion have to be expected. In Figure 1b it can be seen that for given κ the shape of the potential varies from a two minima to a one minimum form as δ grows from 0 to $\delta > \delta_c$. The transition takes place at a critical value $\delta = \delta_c$ given by

$$\cosh^2 [\delta_c - \sqrt{\kappa(\kappa-1)}\,] = \kappa \qquad (41)$$

for which the left hand minimum of the potential disappears at

$$x_c = - \sqrt{\kappa(\kappa-1)} \qquad (42)$$

Because of the symmetry of the potential

$$V(-|x|, -|\delta|, \kappa) = V(|x|, |\delta|, \kappa), \qquad (43)$$

a similar transition takes place for $\delta = -|\delta|$ for which the right hand miminum disappears at $x = |x_c|$. If $x(t)$ is considered to start at the left minimum of $V(x;\delta<\delta_c,\kappa)$ - i.e. with a rationalizing bias - and if $\delta(t)$ moves from $\delta<\delta_c$ to values $\delta>\delta_c$ so that the left minimum of V disappears, $x(t)$ will quickly swing to the remaining right-hand minimum of $V(x; \delta>\delta_c,\kappa)$ at $x>0$ (expansionary bias). This in turn can induce the motion of the alternator $\delta(t)$ to negative values $\delta(t) < -|\delta_c|$, leading to a sudden downswing of $x(t)$ towards a rationalizing bias, and so on. It can be seen that the coupled dynamics of $x(t)$ and $\delta(t)$ can, in principle, explain the upswings and downswings of $x(t)$. In the next section the function of a time dependent alternator shall be explained in economic terms, and its equation of motion will be established.

3. THE EQUATION OF MOTION FOR THE INVESTORS' PROPENSITIES

The alternator δ, which is the investors' strategic choice parameter in the force $K(x;\delta,\kappa)$, and the variable for which an equation of motion is going to be introduced, plays the role of a "trend setting function".

If most investors tend to maximize profits at a given point in time by expanding their business operations so that $x(t)>0$, then some innovators or pioneers (trend–setters) will try to improve their market position by adopting a non-conformist strategy in an attempt to capture quasi-rents due to differentiation. When an upswing is well under way due to expansionary investments undertaken by a majority of investors, these trend setters tend to redirect their effort and start pushing back the cost frontier by means of cost reducing investments. They thereby force others to imitate and also to undertake rationalizing investments in the expectation of further cost reductions along the progress function, descriptive of the least cost combination in the branch of industry under observation.

At other times when a downswing is well under way due to the contractionary effects of rationalizing investments undertaken by a majority of investors, the trend setters start moving towards the quality section of the best practice frontier, introducing better products and implementing investment plans for expanding their facilities. Thereby, others are forced to imitate this expansionary and quality updating behaviour, thus creating the synchronization to be observed in the occurrence of business cycles. Since Adam Smith, economists have viewed these entrepreneurial actions as one of the main sources of the wealth of nations. The equation of motion for the alternator $\delta(t)$, which describes the differentiation activities of entrepreneurs in various fields of industrial investment in aggregate terms, should generate switches under the circumstances stipulated above. A suitable specification of this dynamic behaviour is

$$d\delta(t)/dt = \mu[\delta_0 - \delta(t)]\ \exp[-\beta x(t)] - \mu[\delta_0 + \delta(t)]\ \exp[\beta x(t)] \qquad (44)$$

with $\mu > 0$, $\beta > 0$ and $\delta_0 > 0$.

The mathematical implication of (44) is such that: for $x = 0$, $\delta(t)$ relaxes towards $\delta = 0$; but for $x(t) > 0$ the rhs of (44) represents a strong restoring force so that $\delta(t)$ moves towards $-\delta_0$. The negative δ in turn leads to a force $K(x; \delta, \kappa)$ – see (37) – driving x to negative values. Alternatively, starting from $x(t) < 0$, the rhs of (44) yields a restoring force for a change in $\delta(t)$ towards $+\delta_0$; the positive δ, in turn, produces a force $K(x; \delta, \kappa)$ driving x to positive values again. It can be seen that (44) correctly describes the alternating dynamics of the strategic choice parameter $\delta(t)$ in terms of a non-conformist reaction to the investors' configuration index $x(t)$.

In order to achieve the alternating effects, parameter β, the <u>trend reversal speed parameter</u>, has to be much larger than unity. The <u>strategic flexibility parameter μ</u> describes the flexibility of the investors in turning their strategies from expansionary to rationalizing, and vice versa, whereas <u>the strategic choice amplitude</u> δ_0 is an operative scaling constant.

Using the definitions of hyperbolic functions, the general form of the equation of motion for δ is obtained from (44) as

$$d\delta(t)/d) = -2\mu\ \delta_0\ \sinh[\beta x(t)] - 2\mu\ \delta(t)\ \cosh[\beta x(t)]\ . \qquad (45)$$

Introducing δ_1, a <u>strategy bias parameter</u>, which is positive if the entire trend period is heavily biased towards expansion (as the 1950's were), or negative when biased towards rationalization (as the 1970's were), (45) can be modified and generalized to a complete equation of motion for the alternator. This is both stable enough to hold the ongoing trend over considerable periods of inertia, and flexible enough to change expeditiously with the investors' propensities:

$$d\delta(t)/dt = -2\mu\ \delta_0\ \sinh[\beta x(t)] - 2\mu[\delta(t) - \delta_1]\ \cosh[\beta x(t)]\ . \qquad (46)$$

Analogous to $K(x;\ \delta,\kappa)$, $L(x;\ \delta,\beta)$ is denoted as the strategy reformulation driving force by writing the equation of motion for the alternator in the form:

$$d\delta(t)/dt = L(x;\ \delta,\beta) \quad (47)$$

with

$$L(x;\ \delta,\beta) = -2\mu[\ \delta_0 \sinh(\beta x) + (\delta - \delta_1)\ \cosh(\beta x)] \quad (48)$$

and denoting the parameters as follows:

β: trend reversal speed parameter,

μ: strategic flexibility parameter,

δ_0: strategic choice amplitude,

δ_1: strategy bias parameter.

4. THE CLOSED SET OF EQUATIONS OF MOTION

Collecting the equations of motion for the investors' configuration index $x(t)$ (38), and the equation of motion for the alternator $\delta(t)$ (47), then introducing scaled time τ as

$$\tau = 2\nu t \quad (49)$$

and γ, a scaled strategic flexibility parameter, as

$$\gamma = \mu/\nu, \quad (50)$$

scaled forms of the equations of motion

$$dx(\tau)/d\tau = \hat{K}(x;\ \delta,\kappa) \quad (51)$$

$$d\delta(t)/d\tau = \hat{L}(x;\ \delta,\beta) \quad (52)$$

are obtained with the (scaled) driving force K as

$$\hat{K}(x;\ \delta,\kappa) = \sinh(\delta+\kappa x) - x\cosh(\delta+\kappa x) \equiv [\tanh(\delta+\kappa x) - x]\cosh(\delta+\kappa x) \qquad (53)$$

and the (scaled) strategy reformulation driving force L as

$$\hat{L}(x;\ \delta,\beta) = -\gamma[\delta_0 \sinh(\beta x) + (\delta-\delta_1)\cosh(\beta x)]$$
$$= -\gamma[\delta_0 \tanh(\beta x) + (\delta-\delta_1)]\cosh(\beta x). \qquad (54)$$

It is worthwhile to note that the central equations of the model (51, 52) are invariant against

$$x \to -x,\quad \delta \to -\delta \quad \text{and} \quad \delta_1 \to -\delta_1. \qquad (55)$$

This invariance condition also shows up in the symmetry condition (43) of the potential $V(x;\ \delta,\kappa)$.

5. STRUCTURAL ANALYSIS OF THE SYSTEM OF EQUATIONS

The first part of the analysis examines how many singular points $\bar{P}(\bar{\delta},\bar{x})$ of the equations of motion exist. Singular points are defined as points in the δ-x plane where the motion comes to rest, i.e. where

$$dx/d\tau|_{\bar{p}} = \hat{K}(\hat{x}\hat{\delta}) = 0 \qquad (56)$$

and

$$d\delta/d\tau|_{\bar{p}} = \hat{L}(\bar{x},\bar{\partial}) = 0. \qquad (57)$$

Inserting (53) and (54) into (56) and (57), respectively, it can be seen that the coordinates of a singular point $\bar{P}(\bar{\delta},\ \bar{x})$ have to satisfy the two transcendental equations simultaneously

$$F_1(\bar{\delta}, \bar{x}) \equiv [\tanh(\bar{\delta} + \kappa\bar{x}) - \bar{x}] = 0 \qquad (58)$$

$$F_2(\bar{\delta}, \bar{x}) \equiv [\delta_0 \tanh(\beta\bar{x}) + (\bar{\delta} - \delta_1)] = 0. \qquad (59)$$

These equations are easily solved graphically: the functions $F_1(\bar{\delta},\bar{x}) = 0$ and $F_2(\bar{\delta},\bar{x})=0$ can be represented in the $\bar{\delta}-\bar{x}$ plane and their intersection points, which are the singular points of the differential equations (51, 52), can be determined. Furthermore, it follows from the definitions that the graph of (58) is the locus of all horizontal - and the graph of (59) the locus of all vertical-fluxlines of the equation (51, 52). Restricting the analysis for the moment to the slightly simplified case of vanishing strategic choice bias, i.e. for $\delta_1 = 0$, Figure 2 shows a typical plot of (59) and three typical plots a), b) and c) of (58). This figure confirms that one, three or even five singular points can exist.

The marginal case b) with three singular points, is characterized by the fact that (58, 59) hold good at the singular points $P_+(\delta_+,x_+)$ and $P_-(\delta_-,x_-) = 0$ and $F(\delta,x) = 0$ agree at P_+ and P_- ; thus the condition

$$(dx/d\delta)^{(F_1)}_{P_\pm} = (dx/d\delta)^{(F_2)}_{P_\pm} \qquad (60)$$

where $(dx/d\,\delta)^{(F_2)}$ and $(dx/d\,\delta)^{(F_2)}$ are the derivatives taken along the curves $F_1(\delta,x) = 0$ and $F_2(\delta,x) = 0$, is also satisfied. A straight-forward evaluation of (60) leads to the equivalent implicit condition

$$\kappa = \delta_0\,\beta[1-(\bar{\delta}/\delta_0)^2] + 1/(1-\bar{x}^2) \qquad (61)$$

which also has to be fulfilled by the parameters κ, δ_0 and β for this marginal case. In (61) the solutions δ_+, x_+ or $\delta_- = \delta_+$, $x_- = -x_+$ of (58, 59) have to be inserted for $\bar{\delta}$ and $\bar{x}$ with $\delta_1 \equiv 0$.

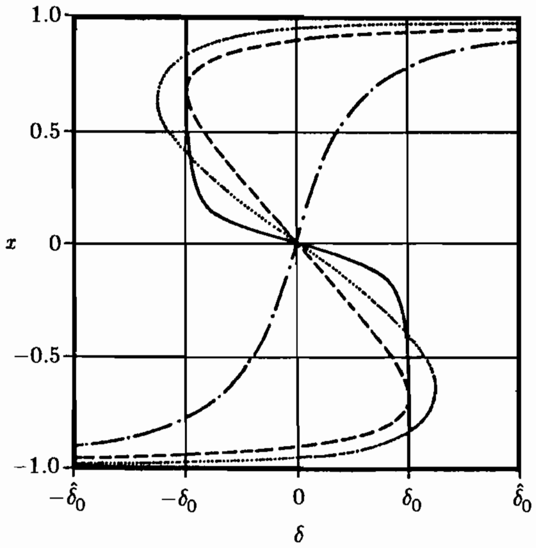


Figure 2 (———) $F_2(\bar{\delta},\bar{x}) = 0$

(— · —) (a) $F_1(\bar{\delta},\bar{x}) = 0$ for small κ (one equilibrium point -- the origin)

(-------) (b) $F_1(\bar{\delta},\bar{x}) = 0$ in the marginal case (3 equilibrium points)

(···········) (c) $F_1(\bar{\delta},\bar{x}) = 0$ for very large κ (5 equilibrium points)

Typical graphs of the transcendental functions

$F_1(\bar{\delta},\bar{x}) = 0$ and $F_2(\bar{\delta},\bar{x}) = 0$

The two-dimensional domain of parameters κ, δ_0 and β satisfying (61), for which there exist three singular points in the δ-x plane, will be denoted by $D_2^{(3)}$. The two-dimensional surface $D_2^{(3)}$ separates the three dimensional domain $D_3^{(5)}$ of the parameters κ, δ_0 β - for which five singular points exist - from the three dimensional domain $D_3^{(1)}$, for which only one singular point (namely the origin $\bar{\delta} = 0$, $\bar{x} = 0$) exists.

To discover the types of paths traversed by $x(\tau)$ and $\delta(\tau)$ it is also relevant to know their behaviour in the vicinity of the singular points. Therefore a linear stability analysis will be performed in order to see whether the singular points are stable or

<u>unstable</u> stationary points. With this objective, the behaviour with time of the solutions in a small neighbourhood of a singular point $\bar{P}(\bar{\delta}, \bar{x})$, is investigated by introducing deviations $\eta(\tau)$ and $\xi(\tau)$ from $\bar{\delta}$ and $\bar{x}$, so that:

$$x(\tau) = \bar{x} + \xi(\tau) \quad \text{and} \quad \delta(\tau) = \bar{\delta} + \eta(\tau). \tag{62}$$

The equation of motion (51, 52) can now be linearized with respect to $\xi(\tau)$ and $\eta(\tau)$. This leads to the following, which are valid only in the vicinity of $\bar{P}(\bar{\delta}, \bar{x})$:

$$d\xi(\tau)/d\tau = a_1\xi + b_1\eta \tag{63}$$

$$d\eta(\tau)/d\tau = -a_2\xi - b_2\eta \tag{64}$$

with the coefficients

$$a_1 = \cosh^{-1}(\bar{\delta} + \kappa\bar{x})\,[\kappa - \cosh^2(\bar{\delta} + \kappa\bar{x})] \tag{65}$$

$$a_2 = \gamma\ \beta\ \delta_0 \cosh^{-1}(\beta\bar{x})$$

$$b_1 = \cosh^{-1}(\bar{\delta} + \kappa\bar{x})$$

$$b_2 = \gamma\ \cosh(\beta\bar{x})$$

where $a_2, b_1, b_2 > 0$, and a_1 is unrestricted. According to standard methods, (63, 64) can be solved using

$$\left.\begin{aligned} \xi(\tau) &= \xi(0)\exp(\lambda\tau) \\ \eta(\tau) &= \eta(0)\exp(\lambda\tau) \end{aligned}\right\} \tag{66}$$

where the eigenvalues λ have to fulfill the determinant condition

$$\begin{vmatrix} a_1-\lambda & b_1 \\ -a_2 & -b_2-\lambda \end{vmatrix} = (a_1-\lambda)(-b_2-\lambda) + a_2b_1 = 0. \tag{67}$$

This quadratic equation yields the eigenvalues

$$\lambda_{1/2} = 1/2[(a_1-b_2) \pm \sqrt{(a_1+b_2)^2 - 4a_2b_1}\,] . \tag{68}$$

The singular point is a <u>stable focus</u> if the eigenvalues λ_1 and λ_2 both have negative real parts, since then both $\xi(\tau)$ and $\eta(\tau)$ approach zero with time. On the other hand, the singular point is <u>unstable</u> if the real part of at least one eigenvalue is positive.

Of great interest is the stability of the singular point $\bar{P}_0(\bar{\delta}_0 = 0, \bar{x}_0 = 0)$ which is, however, slightly shifted from the origin for a non-vanishing strategic choice bias δ_1. With $\delta_1 = 0$, $\bar{\delta} = 0$ and $\bar{x} = 0$, the parameters a_1, a_2 and b_2 assume the simple forms:

$$a_1 = \kappa - 1, \qquad b_1 = 1,$$

$$a_2 = \gamma\beta\delta_0, \qquad b_2 = \gamma, \tag{69}$$

which leads to the eigenvalues

$$\lambda_{1/2} = 1/2[(\kappa-1-\gamma) \pm \sqrt{(\kappa-1+\gamma)^2 - 4\gamma\beta\delta_0}\,] . \tag{70}$$

It is reasonable – and confirmed by the calculations presented in section 6 – to assume relative large values for the trend reversal parameter β and the strategic choice amplitude δ_0. Thus

$$(\kappa-1+\gamma^2) < 4\gamma\beta\delta_0 \tag{71}$$

can be assumed. In this case the eigenvalues (70) become conjugate-complex and can be written as

$$\lambda_{1/2} = 1/2[(\kappa-1-\gamma) \pm i\sqrt{4\gamma\beta\delta_0 - (\kappa-1+\gamma)^2}\,] . \tag{72}$$

From (72) it follows immediately that a stable focus exists at the origin of the δ-x plane under the condition

$$\kappa-1-\gamma < 0 \quad \text{or} \quad \kappa < 1+\gamma \tag{73}$$

and that the origin becomes an unstable focus if condition

$$\kappa-1-\gamma > 0 \quad \text{or} \quad \kappa > 1+\gamma \tag{74}$$

is fulfilled.

Finally, the solutions of the equations of motion (51, 52) have to be investigated. A complete survey of all the possible types of solution to the equations dependent on the parameters $\kappa, \delta_0, \delta_1, \beta$ and γ will not, however, be made. The immediate aim is to derive sufficient conditions for those parameters which lead to the existence of an asymptotically periodic type of solution; this is the type of solution which has been anticipated qualitatively in terms of the investment cycle, and in which the main interest lies. In other words, it is necessary for the substantial aspects of the model being considered to derive an existence theorem for solutions approaching a "limit cycle" under certain conditions of the parameters $\kappa, \delta_0, \delta_1, \beta$ and γ.

A limit cycle C(t) is defined as a closed trajectory, i.e. a periodic solution to the equations of motion, with the property that there exists a domain D_c around C(t), so that all trajectories starting within D_c approach C(t) as $t \to \infty$. D_c can be denoted as the "domain of attraction"' and C(t) as an "attractor". Before the existence theorem is stated, the famous Poincaré-Bendixon theorem should be formulated in a – in context – relevant version:

Consider two autonomous first order differential equations for the variables $\delta(\tau)$ and $x(\tau)$, and suppose that a finite domain D_c exists in the δ-x plane such that:

a) no singular points are situated in D_c

b) all trajectories $\delta(\tau)$, $x(\tau)$ starting inside or on the boundary of D_c at time $\tau = 0$ remain in D_c for $0 < \tau < \infty$. In this case (at least one) limit cycle must exist within D_c

and all trajectories in D_c either <u>are</u>, or <u>approach</u> as $\tau \to \infty$ a limit cycle.

Suppose that the parameters $\kappa, \delta_0, \beta, \gamma, \delta_1 = 0$ satisfy the conditions

1) $\{\kappa, \delta_0, \beta\} \in D_3^{(1)}$.
   i.e. the origin is the only singular point (75)
2) $(\kappa-1+\gamma)^2 < 4\gamma\beta\delta_0$
3) $\kappa > 1 + \gamma$.

These imply that the eigenvalues λ_1 and λ_2 belonging to $\bar{P}(\bar{\delta} = 0, \bar{x} = 0)$ are conjugate-complex with a positive real part, so the one singular point P(0,0) is an unstable focus. D_c, of the δ-x plane is bounded internally by an infinitesimally small elliptical core enclosing the origin and externally by the straight lines $\delta = \delta_0$, x = 1, $\delta = -\delta_0$ and x = -1 (Figure 3). It fulfills the premise of the Poincaré-Bendixon theorem that: a) there are no singular points situated in D_c, and b) all trajectories starting from the boundary of D_c enter D_c and remain in D_c.

Premise a) follows from condition 1) of (75), according to which the origin is the only singular point.

Premise b) follows from the fact that according to conditions 2) and 3) of (75), the origin is an unstable focus with all trajectories spiraling out of it and therefore entering the core from its interior boundary C_i. Also, on the exterior boundary C_e:

$$dx/d\tau = [\tanh(\delta+\kappa) -1] \cosh(\delta+\kappa) < 0 \quad \text{for} \quad x = 1$$
$$dx/d\tau = [\tanh(\delta-\kappa) +1] \cosh(\delta-\kappa) > 0 \quad \text{for} \quad x = -1 \tag{76}$$

$$d\delta/d\tau = -\gamma[\delta_0 \tanh(\beta x) + \delta_0] \cosh(\beta x) > 0 \quad \text{for} \quad \delta = \delta_0$$
$$d\delta/d\tau = -\gamma[\delta_0 \tanh(\beta x) - \delta_0] \cosh(\beta x) > 0 \quad \text{for} \quad \delta = \delta_0 \tag{77}$$

and (76, 77) show that all trajectories crossing C_e are directed inward into D_c. From the Poincaré-Bendixon theorem, it now follows that within a limit cycle, i.e. a periodic solution of the equations of motion, C(t) must exist, and that all other solutions starting from any point within D_c approach this limit cycle. D_c is a domain of attraction for C(t). Assumptions 1-3) are sufficient but not necessary conditions for the existence of a limit cycle. This means that there could also exist limit cycle solutions under more general conditions for the parameters κ, δ_0, δ_1, β and γ. On the other hand, it can be shown that there are initially quasi-cyclic trajectories which break down into a stable solution after some oscillations.

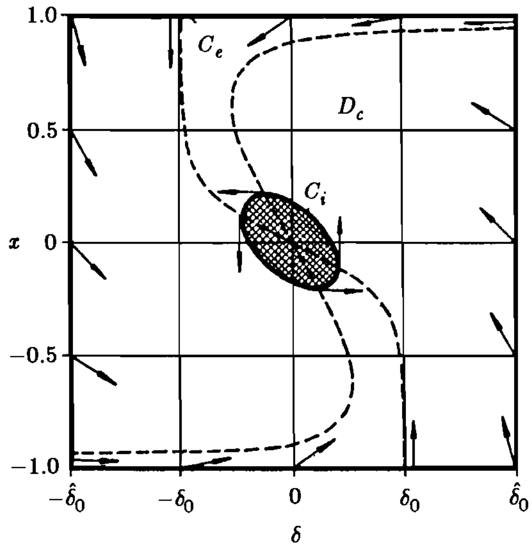


Figure 3 Domain D_c with interior boundary C_i and exterior boundary C_e. All trajectories enter D_c and remain in D_c as the flux lines at the boundaries show.

If assumption 3) of the existence theorem is changed into

$$\kappa < 1 + \gamma \tag{78}$$

the case for which the origin is the only singular point and is a stable focus is obtained. The main importance of these two different types of solution lies in the implied proof that whether or not an economy tends towards a stationary state, or undergoes non-equilibrium oscillations, depends in a sensitive way on the decision psychology of the entrepreneurs as expressed by the decision psychology parameters exist. κ, δ_0, δ_1, β and γ. In particular "critical values" of these parameters exist. Crossing these critical thresholds elicits a transition from one type of solution into the other. (See $\kappa = 1.6$, which crosses the marginal case with a critical threshold value of the coordinator $\kappa_c = 1.5$).

In Figure 4 a parameter combination is chosen for which the origin $\bar{\delta} = 0$, $\bar{x} = 0$ is the only singular point, and a stable focus of the damping out of cyclic behaviour. This phenomenon which troubles disequilibrium models in situations of zero-growth, is but a special case among many others in the non-equilibrium model being discussed.

Sustained cycles occur for a parameter combination which satisfies the assumptions of the limit cycle existence theorem. Figures 5a and 5b exhibit the limit cycle. Again, the origin is the only singular point which, however, has now become an unstable focus of the motion of the economy.

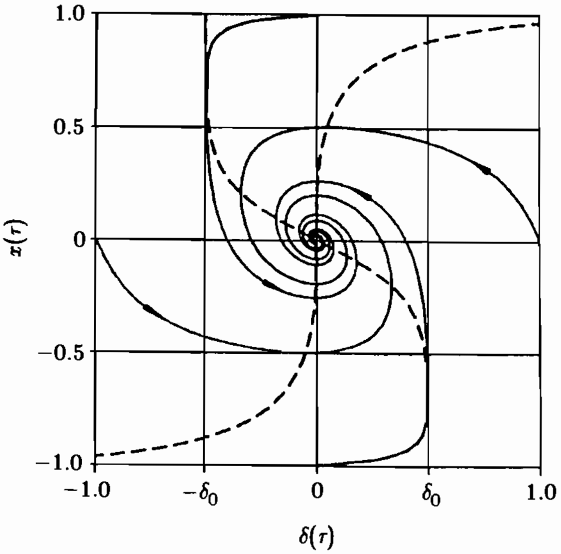


Figure 4 (a) Trajectories in the δ-x plane approaching the one stable focus at the origin $P_c(0,0)$ for κ= 1.0, δ_0= 0, β = 4.0 and μ = 0.5.
(———) Trajectories; (------) $F_{1,2}(\bar{\delta},\bar{x}) = 0$

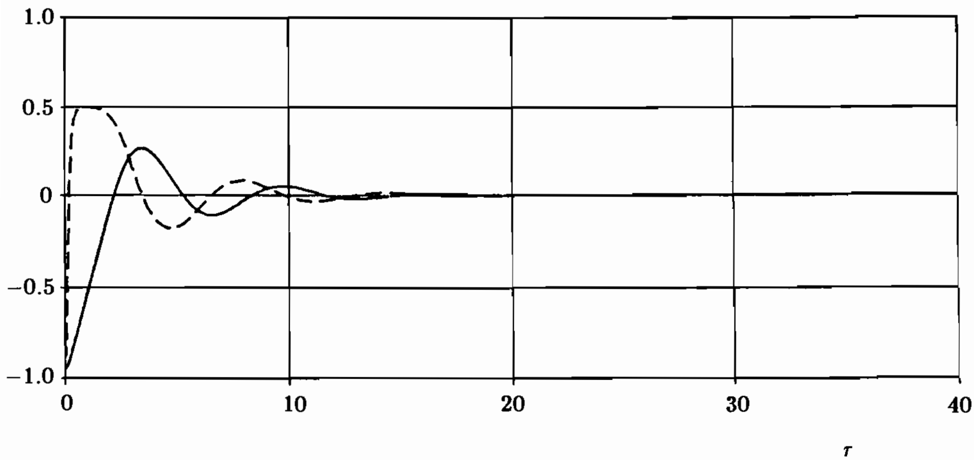


Figure 4 (b) Path of x(τ) (———) and δ(τ) (------) for the parameters of 3.4a)

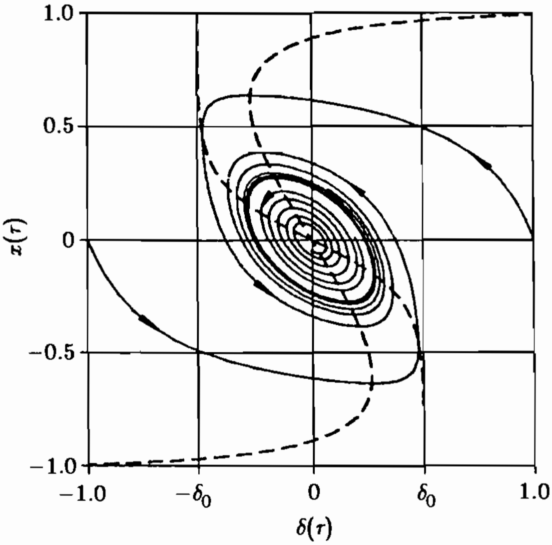


Figure 5 (a) Trajectories in the δ-x plane for one unstable focus at $P_e(0,0)$ and one limit cycle for $\kappa = 1.6$, $\delta_0 = 0.5$, $\delta_1 = 0$, $\beta = 4$ and $\mu = 0.5$. These fulfill the condition (75) of the limit cycle existence theorem. (————) Trajectories: (------) $F_{1,2}(\bar{\delta},\bar{x}) = 0$.

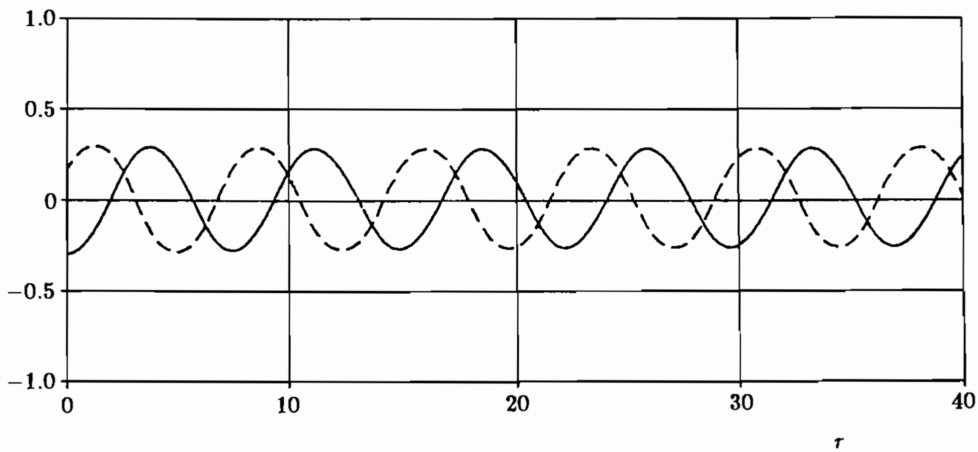


Figure 5 (b) Path of $x(\tau)$ (————) and $\delta(\tau)$ (------) for the parameters of (a). The motion is periodic and traverses the limit cycle.

6. CHANGES IN INDUSTRIAL STRATEGIC INVESTMENT IN THE FEDERAL REPUBLIC OF GERMANY BETWEEN 1956 AND 1978

It has just been demonstrated numerically by means of parametrical variation of some of the influential factors in the model, that theoretical solutions can represent a whole range of types of economic motion. A variety of periodic, symmetric and asymmetric oscillations, and also some vanishing, accelerated and decelerated fluctuations that approach an existing mean value trend line, or diverge towards a new, higher or lower trend line, have been demonstrated. And it is exactly this variety of possible types of motion which has to be explained when theory is compared with the variety of economic motions observable empirically. It has therefore been shown that by incorporating the micro-economic strategic investment decisions of entrepreneurs into the model, macro-economic fluctuations can be interpreted, and that the influence of macro-economic change on the micro-economic level of entrepreneurship can be explained.

In devising a test for the model, with data for a real economy, two restrictions apply. The first restriction is spatial. The non-availability of time series on industrial strategic investments (expansionary and rationalizing) in some countries, narrows down the choice of national economies to which the model can be readily applied. But even in countries such as United States of America - where data are available, but only for one type of strategic investment (expansionary) and not for the other rationalizing type (the latter is reported with tactical investments or replacement) - some relevant empirical regularities could be predicted using the model. For example, modernization expenditures (Feldstein and Foot 1971) vary over time, while replacement itself may be regarded as a constant proportion of capital stock: (Jorgenson (1963), Jorgenson and Stephanson (1967)). On the other hand, the fluctuations in replacement and modernization expenditures are less pronounced than the fluctuations in expansionary investment, which, according to Eisner (1978), explain almost 80% of the variation in annual capital expenditures in the United States' business sector. Consequently, it can be predicted, using statistical methods, that the explanatory power of an investment analysis falls under the 80% line suggested by Eisner if the analysis mixes expansionary investment with investment for modernization and replacement, since the analysis of "strategic" investment would be "contamined" with some "tactical" investment, namely for replacements.

Conversely, the explanatory power of an analysis of industrial investment in plant and equipment in the United States would probably exceed 80% if the analysis were restricted to "strategic" investment proper, i.e. expansionary and rationalizing investment (the latter now being buried under "modernization and replacement"). The availability of pertinent data is one of the reasons for the choice of the Federal Republic of Germany as the test unit for analysis.

The second restriction upon the application of the model as it stands in temporal. Times of discontinuity have been excluded from the test as these represent changes in economic regime, or phase transitions. Although such transitions can be handled by the model, additional theory (long term economic theory) would have to be incorporated – for instance, in the form of an equation of motion for the coordinator – in order to handle discontinuities systematically. It is well known that differential equation models of only three coupled dynamic variables (e.g. the Lorenz model) already possess "strange attractor" solutions side by side with those of the limit cycle type. Though fully deterministic, such solutions could represent quasi-chaotic economic motion. These questions of long term economic theory and mode architecture, and how they relate to possible phase transitions in times of discontinuity, will not be discussed further here (see, however, Mensch, Weidlich, Haag, 1987).

For the period 1955-1980 data are available on expansionary and rationalizing industrial investment in the Federal Republic of Germany, and moreover, this period appears to be void of a major discontinuity. Nevertheless, this period certainly does not build a continuum. In terms of economic policy and government regulation, it can be divided into two main sub-periods each of about eleven years. The first four or five years of the second sub-period can clearly be designated as "interventionistic". Investment activity was therefore under the influence of the "regulator" δ_1 to varying degrees, differing during the three periods 1955-1965, 1967-1971 and 1973-1980. The increasing $g(t)$ mainly reflects the general increase in the total number of investment projects. Furthermore, the technological base of German industry changed and matured during these years, in more or less the way suggested and modelled by Utterback and Abernathy (1975), as well as Mensch (1979), and Mensch, Kaasch, Kleinknecht, Schnopp (1980).

According to their findings, industries develop in terms of technology and market structure, by going through stages which depend on the age of the technology and other related factors in the particular industry concerned. The following regularity holds for the post-war growth industries in most western countries: the 1950s were dominated by product innovation in diversifying industry: the 1960s by process innovation in concentrating industry, and the 1970s by pseudo-innovations in nearly all industries; the 1980s are expected to be dominated by a cluster of basic innovations in some key industries. Therefore it can be maintained that these patterns in the rate and direction of innovation have created, and have been the consequence of, a high degree of conformity in strategic behaviour (namely, in the timing of innovation, synchronization of investment and imitation of business tactics) in the first sub-period under consideration. In the second sub-period, on the contrary, there appears to have been a low degree of conformity. This alteration is reflected in the relative values of the "coordinator" κ chosen in modelling the three periods . Similarly, in the three periods the value of the "accelerator" ν was chosen to be "small", "medium" and "large", because it reflected the shortening of successive product life cycles as major industries advanced out of a phase of large improvement effects (true product innovation) to minor improvement effects (mere pseudo-innovation) which did not last very long.

In Table 1 the chosen values of the parameters are listed for the reader, who may wish to replicate the empirical analysis that follows; the investment data for the Federal Republic of Germany is given in Table 2.

Table 1 Model Parameters for the Economy of the Federal Republic of Germany

Period	γ	β	δ_0	δ_1	g	κ	ν
1955-1965	0.13	10	0.5	-0.3	0.2	1.5	0.145
1967-1971	0.13	10	0.5	+0.3	0.5	1.5	0.225
1973-1980	0.13	10	0.5	-0.3	0.7	0.7	0.400

The original data were collected by a German institute for economic research (The IFO-Institute in Munich) by means of questionnaires sent to a (representative) sample of German industrial corporations. They have been econometrically corrected for a stronger rationalization bias (Mensch et.al. 1980), so that the regression line of Figure 6a, which depicts the data, possesses a smaller intercept and a steeper descent than it would have in the original data version, which incorporates an expansionary bias. Figure 6b shows the investment structure data z(t), which is the investment structure index Z(t) cleaned of the linear trend. Weidlich and Haag (1982) have shown that the trend deviation pattern is robust, i.e. insensitive to the form of the trend functions. Also, since the investment structure index without trend is to be used, the difference in built-in biases between the original data version and the Mensch version should have no effect on the matching of the observed trend deviations with the theoretical values calculated from the model.

<u>Table 2</u> Investment Data for the Federal Republic of Germany 1956-78

Year	R-investment [10^9 DM]	E-Investment [10^9 DM]	Z = (E-R)/(E+R)
1956	16.4	11.4	-0.183
1957	15.9	10.8	-0.197
1958	16.6	8.2	-0.342
1959	17.7	10.2	-0.274
1960	22.2	12.6	-0.282
1961	24.2	13.6	-0.282
1962	24.2	11.2	-.0369
1963	22.4	10.6	-0.360
1964	23.5	11.2	-0.360
1965	26.1	13.0	-0.342
1966	25.9	11.6	-0.388
1967	23.0	7.2	-0.527
1968	22.3	10.4	-0.369
1969	30.0	21.6	-0.162
1970	34.9	23.2	-0.205
1971	34.4	18.6	-0.298
1972	31.3	12.4	-0.439
1973	30.5	13.8	-0.379
1974	28.5	10.2	-0.471
1975	25.9	6.2	-0.626
1976	26.5	6.8	-0.626
1977	27.5	6.6	-0.613
1978	28.9	7.0	-0.613

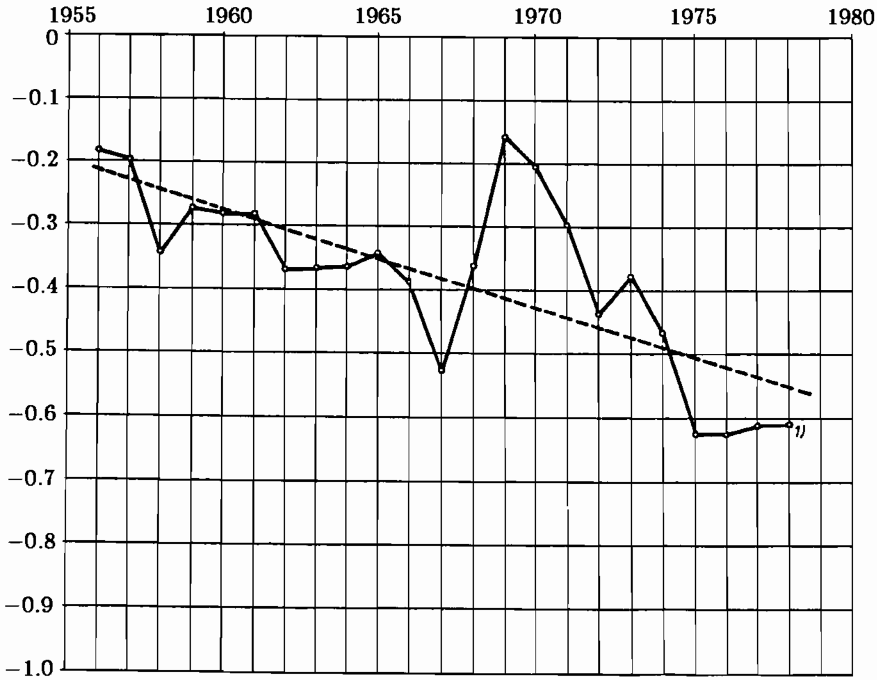


Figure 6a Investment structure index $Z(t) = Z_0 + z(t)$ in the Federal Republic of Germany 1956-1978 (data by IFO). A linear time dependence of the long-term component $Z_0(t)$ (------) is assumed.

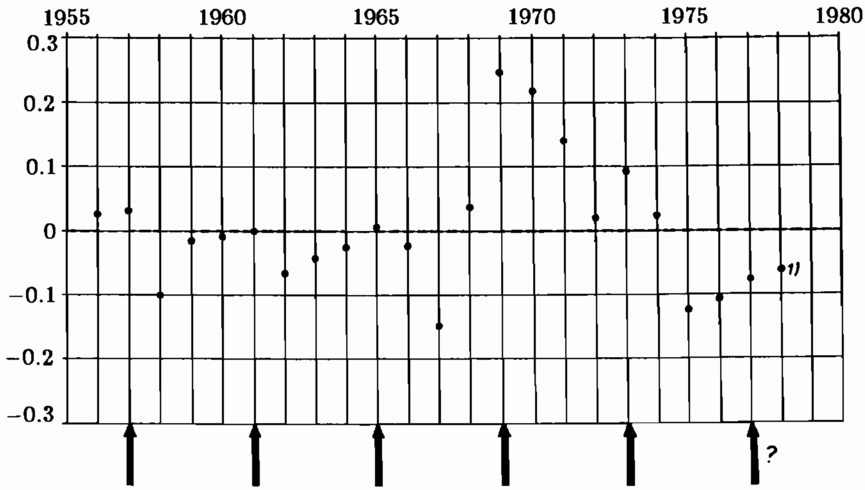


Figure 6b Short-term investment structure index z(t).
The long-term trend component $Z_0(t)$ of Figure 6a has been subtracted.

Figure 7 shows the result of the empirical analysis, which consisted of a piecewise application of the model to the set of parameters specified in Table 1, for the three periods 1955-1971, 1967-1971 and 1973-1980.

From the mid-fifties through the mid-sixties the matching external conditions chosen correspond to those implying the existence of a limit cycle as indicated by the limit cycle existence theorem. During these years, the Federal Republic was still enjoying its "economic miracle", which was nurtured by both post-war reconstruction of plant, equipment and infrastructure, and by a high rate of product innovation in a number of growing international industries. A strong expansionary bias and bias reinforcement (κ = 1.5), however, led to a strongly non-sinusoidal distortion of the periodic motion of the investment structure index. Again, as in the case of the United States, the fluctuation in expansionary investment explains most of the variation in the economic indicators of aggregate change. Even in the expansionary phase, industrial investors changed quickly to rationalizing investment after a relatively short expansionary boom, and they only reverted slowly in the direction of expansionary investment, even after a relatively long period of rationalization. Thus, the effects of entrepreneurial decision processes, as reflected by the path of $z(t)$ in Figure 7, reveal a persistent, and perhaps growing concern for rationalization and productivity advancement (even in the period 1955-65 when the economy expanded at a high rate). On the other hand, preparations for the "quantum leaps" which occurred were time-consuming. The development of new or better products, and the planning of new or bigger plants to produce them, took place during the preceding periods of growth in the business cycle.

During the period 1955-65, the strategic choice parameter $\delta(t)$, the "alternator", stayed only briefly in the rationalization mode and switched back quickly into the expansionary mode as soon as rationalization investment strategy had taken effect. From the mid-sixties onwards, however, the alternator took a different course, staying mostly in the rationalization mode, and only rarely and briefly switching over to the expansionary mode. The drop in the propensity to finance extended plant and equipment, as well as the research and development work which leads to growing business, is seen to occur as early as 1966 to 1967. As process innovation was the gist of the rationalization atmosphere that prevailed from about 1965 on, a confirmation

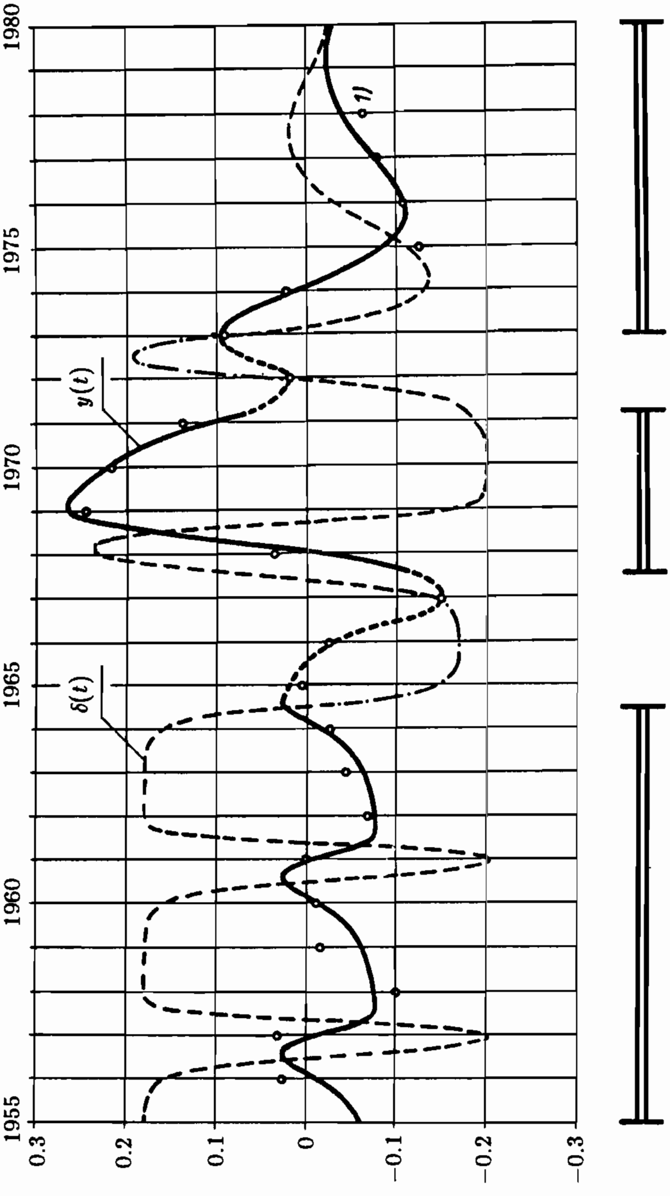


Figure 7 Path of the short-term component z(t) of the investment structure index (———) and of the alternator δ(t) (------) for appropriate choices of trend parameters in comparison with the observed investment structure data, West German Industry 1956-1978.

of the Utterback-Abernathy model of product and process innovation has been obtained with the use of German statistics. This indicates a fairly general switch-over from product innovation to process innovation at about that time. Furthermore, as already suggested by Mensch (1979), the seventies was a period of relatively weak propensity to innovate both products and processing in comparison to the previous years, and the time-path of the alternator found to fit the observed investment behaviour does indeed portray the predominance of pseudo-innovation during the 1970s.

During the late sixties , after the neo-liberal incumbant government fell, the newly formed "Grand Coalition" in the Federal Republic of Germany resorted to a massive expansionary policy programme. Its effect on the investment structure index is clearly visible as an all-time peak in 1969/1970, called "Schiller Effect", after Schiller, the accomplished Keynes economist who became both minister of finance and of economic affairs. This irregularity in economic conditions has been adjusted for by temporarily boosting the "regulator" δ_1. After the Schiller period, the regulator had to be returned to its initial level in ord^er to achieve the fit depicted in Figure 7 for the period after 1972. This may indicate more than the simple necessity to alter the regulator in order to model unusual periods. It may lend further evidence to the not unusual view that forceful demand management only forces industrial firms to consider the supply side limits to growth more quickly.

Although some economists argue that the oil shock caused the "slumpflation" after the hyperboom around 1970 (Freeman, Soete and Townsend (1982)), an evolutionary explanation may be more appropriate. For the period after 1971, it was not possible to fit the observed motion of the investment structure index (net of trend) with model data on the basis of parameters which satisfy the conditions of the limit cycle existence theorem. The matching values of the "inflator" g - which could not be set lower with the given price developments - and of the "accelerator" ν - which reflects the short-ended product life cycles of pseudo-innovations - fit together only by using a "coordinator" value κ below unity, which is in the realm of damped fluctuations. Marketing managers speak of "a fluttering cycle" in reference to this quasi-stationary state, in which the low amplitude and high frequency of the branch cycles do not create either an upswing or a true downswing in macro-economic aggregate.

Obviously this is a state of high micro-economic friction and loss of activity, which already carries the seeds for a transition into another cyclic stage in which the investors' interaction parameter κ would exceed the critical value κ. Needless to say, the economy need not necessarily resume its cyclic motion with an upswing from the current intermediate level of investment activities.

This model is not designed to predict the investment structure index. But it can be used for diagnostic purposes. If the data are meaningful and the model is reasonable, (as the authors believe), then industry in the Federal Republic is in a state of de-phasing (asynchronization of micro-economic developments for the diverse firms, industrial branches and sectors). Thus, it is getting "ready" for a phase transition, and preparing itself structurally for the breakthrough of a whole set of basic innovations. Whether this will in fact take place, could only be answered by further investigations which encompass many more factors than the short-term business cycle theory presented here.

REFERENCES

Barzel, Y., 'Optimal Timing of Innovations', Rev. Economics and Statistics 50, 1968, pp. 348-355.

Chamberlin, E.H., Toward a More General Theory of Value, Oxford University Press, New York, 1957.

Chang, W. and Smyth, D., 'The Existence of Persistence of Cycles in a Nonlinear Model, Kaldor's 1940 Model Re-Examined', Rev. Economics and Statistics 32, 1950.

Eisner, R., 'Factors in Business Investment', National Bureau of Economic Research, General Series NO. 102, Ballinger, Cambridge, MA, 1978.

Feldstein, M.S., and Foot, D.K., 'The Other Half of Gross Investment: Replacement and Modernization Expenditures', Rev. Economics and Statistics 53, 1971, pp. 49-58.

Flaherty, M.T., 'Industry Structure and Cost-reducing Investment', Econometrica, 48, 1980, pp. 1187-1209.

Freeman, C., Soete, L., and Towsend, J., Fluctuations in the Numbers of Product and Process Innovations, mimeographed paper, Brighton, 1982.

Goodwin, R.M., ' A Non-Linear Theory of the Cycle', Rev. Economics and Statistics 32, 1950.

Goodwin, R.M., "The Nonlinear Accelerator and the Persistence of Business Cycles', Econometrica 19, 1951, pp. 1-17.

Glaser, O. and Strauss, L., 'Awareness Contexts and Social Interaction', Am. Sociological Rev., 29, 1964, pp. 669-679.

Hicks, J.R., A contribution to the Theory of the Business Cycles, 3rd ed., Clarendon Press, Oxford, 1965.

Jorgenson, D.W., 'Capital Theory and Investment Behaviour', Am. Economic Rev. 53, 1963, pp. 241-259.

Jorgenson, D.W. and Stephenson, J.A., 'Investment Behaviour in U.S. Manufacturing: 1947-1960', Econometrica 35, 1967, pp. 169-220.
Kalecki, M., 'A Macrodynamic Theory of Business Cycles, Econometrica 3, 1935, pp. 327-348.
Kamien, M.L. and Schwarz, N.L., 'Timing of Innovation Under Rivalry', Econometrica 40, 1972, pp. 43-60.
Keynes, J.M., 'The General Theory and After, Part II: Defence and Development' in Moggeridge, D. (ed.) Keynes Collective Writings, vol. 14, London 121, 1973.
Mensch, G., Stalemate in Technology, Ballinger, Cambridge, MA., 1979.
Mensch, G., Kaasch, K, Kleinknecht, A., and Schnopp, R., Innovation, Investment Trends and Switches Between Full- and Underemployment Equilibria, Berlin, International Institute of Management, Discussion Paper 10 (1980).
Mensch, G., Weidlich, W., Haag, G., Outline for a Formal Theory of Long-Term Economic Cycles, in (ed.) Vasko, T., The Long Wave Debate: Concepts, Results and Perspectives, IIASA-publication, 1987.
Rosenberg, N., 'On Technological Expectations', Economic J. 86, 1976, pp. 523-535.
Salter, W..G., Productivity and Technical Change, Ballinger, Cambridge, MA, 1960.
Schumpeter, J.A., Konjunkturzyklen eine theoretische, historische und statistische Analyse des kapitalistischen Prozesses, vol. 2, Vandenhoeck und Ruprecht, Göttingen, 1961.
Utterback, J.M. and Abernathy, W.J., 'A Dynamic Model of Process and Product Innovation', Omega., Intern. J. Management Sci. 3, 1975, pp. 639-654.
Weidlich, W. and Haag, G., An Indicator of Change in the Investment Structure, to appear in Proceedings of the Berlin Conference on "Modelling Innovation Processes and Structural Change" ed. by G. Mensch, International Institute of Management, Science Center, Berlin, 1982.
Weidlich, W. and Haag, G., Concepts and Models of a Quantitative Sociology – The Dynamics of Interacting Populations, Springer Series in Synergetics, vol. 14, Berlin, 1983.

Chapter 10

Complex Dynamics in Continuous Models of the Business Cycle

T. PUU

1. INTRODUCTION

The golden age of economic dynamics was the 1950s, when so many models – in the form of difference, differential, and integral equations – were formulated. The modelled processes were essentially nonlinear. Unfortunately, the results for the nonlinear models were obtained by outright calculation with the means then available, or – as it is nowadays called – by simulation. Of course, only small sets of initial conditions could be dealt with in that manner. Accordingly, the results had a more or less ad hoc character.

By contrast, the physical theory for linear oscillators and its well developed mathematical tools provided a rich body of systematic knowledge, and the temptation to work with linear accelerators and multipliers must have been great. After all, the method of linearisation had worked well in physics and it was satisfying to obtain explicit solutions in the shape of sinusoid motion or exponential growth.

However, there is a methodological snag in this. Linearisation is a first approximation that can remain valid only in sufficiently small intervals for the variables. Accordingly, the complete model cannot itself produce changes that force the variables outside these intervals, or it will destroy its own assumptions. This is precisely the case with explosive cycles and exponential growth. Moreover, standing cycles or zero growth in linear models can only occur when the parameters stand in a very specific relation in a structurally unstable model, bound to change qualitatively at the slightest perturbation (see Arnold, 1983).

Accordingly, the only thing that linear models can decently produce is a progressive decay of any initial movement and a return to eternal equilibrium. One can then ask in what sense the model is dynamic, as it is only capable of modelling the decay of change introduced exogeneously. These facts must have been obvious to Hicks (1950) when he introduced the concepts of "floor" and "ceiling", and to Goodwin (1951) when he made the accelerator nonlinear. These, unfortunately, were exceptions to the mainstream of linear modelling.

Hicks combined numerical results from calculations using his difference equation model with intuitive reasoning of great insight. Goodwin, working with a continuous time model, actually employed the Poincaré-Bendixon theorem to demonstrate the existence of a limit cycle.

The perturbation methods - developed in the 1950s and 1960s to deal with nonlinear oscillatory systems when computation capacity was severely limited - would have been useful for models of the Hicks-Goodwin type, but they never diffused to the economics profession. Concerning these methods see Stoker (1950), Hayashi (1964), Jordan and Smith (1977), and Kevorkian and Cole (1981).

Today the whole field of topological dynamics has experienced an explosive development and fascinating phenomena such as bifurcations, catastrophes, strange attractors, turbulence, and chaos have been studied. Some of these concepts have disseminanated through the general public, such as the ignenious four volume set of non-mathematical "cartoons" by Abraham and Shaw (1982-85).

Reconsidering economic macrodynamics in this new light seems to be an interesting task, particularly in the area of muliplier-accelerator models of the business cycle. However, there still remains a choice of representation in discrete or in continuous time. In general, discrete modelling is needed whenever actual computation or empirical implementation is involved, whereas continuous formulations tend to be superior for general theoretical reasoning. Primarily, however, the choice is one of convenience. It seems to have been purely accidental that growth was modelled in continuous time and business cycles in discrete time. Accordingly, we should feel free to choose the most convenient representation.

A useful continuous representation of a multiplier-accelerator process is the one due to Phillips (1954). We take its formulation from Allen (1956). Denote income by Y, investments by I, the propensity to save by s and the accelerator by v. Denote any so called "autonomous" expenditures by A. Then we have

$$\dot{Y} = \lambda(A + I - sY), \tag{1}$$

$$\dot{I} = \kappa(v\dot{Y} - I) \tag{2}$$

where λ and κ denote two adjustment speeds. Provided these speeds are infinite and A is constant, we obtain the Harrod (1948) model of balanced growth with its familiar solution $Y = A/s + (Y_0 - A/s) \exp[(s/v)t]$.

With finite adjustment speeds the appropriate procedure is to differentiate (1) once more and then substitute from (1)-(2) to eliminate investments. In this way

$$\ddot{Y} + \lambda\kappa sY = (\lambda\kappa v - \kappa - \lambda s)\dot{Y} + \lambda\kappa A + \lambda\dot{A} \tag{3}$$

is obtained. Depending on the sign of the coefficient of the $\dot{Y}$ term, the solution of the homogeneous equation – obtained by deleting the autonomous terms – is an explosive or damped harmonic oscillation with the exponential rate of damping $\alpha = (\lambda\kappa v - \kappa - \lambda s)/2$ and the frequency $\omega = \sqrt{(\lambda\kappa s - \alpha^2)}$. Due to the principle of superposition, the solution may be added to any particular solution for (3). When A is independent of time, the constant value A/s is such a particular solution.

2. NONLINEAR INVESTMENT FUNCTION

Hicks (1950) assumed that the linear accelerator $v\dot{Y}$ is in action until the process hits either of two linear constraints: the floor, when income decreases faster than capital due to natural depreciation; or the ceiling, when income increases faster than other essential inputs, which, unlike capital, are not endogeneous to the model. Goodwin (1951) assumed that the response of investments to income changes, being governed by the accelerator around the origin, tapers off asymptotically with large income changes.

A formalization would be to replace $v\dot{Y}$ by $vf(\dot{Y})$ where $f(0) = 0$, $f'(0) = 1$ and $\text{Lim } f(\dot{Y}) = +a, -b$ as $\dot{Y} \rightarrow \pm \infty$. A simple smooth function with these properties, employed by the author in Puu (1986), is $\tanh(\dot{Y})$, where $a = b = 1$, or its truncated Taylor series $\dot{Y} - \dot{Y}^3/3$.

We shall temporarily disregard the autonomous expenditures and record the resulting nonlinear equation in its homogeneous form. In passing, note that as long as the nonlinearity is confined to the $\dot{Y}$ term alone, the superposition principle still holds. The homogeneous equation

$$\ddot{Y} + \lambda\kappa s Y = \lambda\kappa v f(\dot{Y}) - (\kappa + \lambda s)\dot{Y} \tag{4}$$

is known as Liénard's equation. It has been shown to possess a limit cycle solution under the assumptions concerning the investment function stated; see Jordan and Smith (1977). There is even a constructive geometrical method for drawing the orbit; see Stoker (1950) or Hayashi (1964). The conclusion remains valid even if the savings function had been nonlinear as well, provided savings increase with increasing income, as is generally assumed.

3. EXISTENCE OF A LIMIT CYCLE

It is instructive to see the reason for the existence of a limit cycle solution. To this end, we multiply (4) by $\dot{Y}$ obtaining

$$\frac{1}{2}\frac{d}{dt}(\dot{Y}^2 + \lambda\kappa s Y^2) = \lambda\kappa v f(\dot{Y})\dot{Y} - (\kappa + \lambda s)\dot{Y}^2. \tag{5}$$

The parenthesis on the left hand side is an energy integral of Lyapunov type. Equation (5) tells us when the system is gaining or losing energy. Suppose that the rate of change of income is very close to zero. If the system is thus approaching its equilibrium income, which, because of the homogeneous form (4) must itself be zero, then the energy must become zero. However, suppose that

$$\lambda\kappa v > \kappa + \lambda s. \tag{6}$$

As we have assumed that $f'(0) = 1$, there is a neighbourhood of the origin where $f(\dot{Y})\dot{Y}$ is approximately equal to $\dot{Y}^2$, and so, according to (6), the right hand side of (5) is positive. The conclusion is that if change is close to extinction then the system is gaining energy, either in terms of an increasing kinetic energy $\dot{Y}^2$, or in terms of an increasing potential energy Y^2, the latter being defined in terms of the deviation from equilibrium.

If, on the other hand, $\dot{Y}$ is very large, either positive or negative, then the first term on the right hand side of (5) is approximately linear in $\dot{Y}$, with $f(\dot{Y})\dot{Y}$ always positive. This is due to the assumption that always $-b < f(\dot{Y}) < a$. The second term, being negative, is still quadratic, and accordingly bound to dominate for $\dot{Y}$ sufficiently large. Accordingly, the system is damped for large rates of change.

The combination of damping at large rates of change and antidamping at small rates of change usually leads to limit cycles. We can deal with the existence problem formally. First, let

$$E = \dot{Y}^2 + \lambda\kappa s Y^2. \tag{7}$$

Next define E sufficiently small for

$$\lambda\kappa v f(\sqrt{E}) > (\kappa + \lambda s)\sqrt{E} \tag{8}$$

to hold. When f is a smooth function with $f\,'(0) = 1$ this can always be done whenever (6) is fulfilled. Consider the points in the $Y,\dot{Y}$ plane satisfying (7) for this value of E. Then the right hand side of (5) is positive, and we conclude from (5) that E is increasing over time. In other words, any trajectory crossing the elliptic boundary defined by (7) is outward and never enters the interior of this ellipse again.

Let us now consider large values of E instead. Using definition (7) in equation (5), and by subtracting εE from both sides we obtain

$$\dot{E}/2 - \varepsilon E = \lambda\kappa v f(\dot{Y})\dot{Y} - (\kappa + \lambda s + \varepsilon)\dot{Y}^2 - \varepsilon\lambda\kappa s Y^2. \tag{9}$$

For a sufficiently large E the right hand side of (9) obviously has to be negative due to the fact the $f(\dot{Y})$ is bounded. Accordingly we get

$$d/dt \ln E < 2\varepsilon \quad (10)$$

for sufficiently large E. Now, ε can be made as small as we wish, and as the limit decreases to zero we see that E has to decrease with time. By choosing ε sufficiently small and E sufficiently large we see that the trajectory does not escape the ellipse defined in the phase plane by (7). Accordingly, there is an elliptic annulus in the phase plane from which no trajectory escapes.

Due to the Poincaré-Bendixon theorem the system then moves to a limit cycle, as there are no equilibrium points within this annulus. For a stringent proof of existence along different lines and for somewhat more specific f-functions the reader is referred to Jordan and Smith (1977).

4. APPROXIMATION OF THE LIMIT CYCLE

In Figure 1, the limit cycle for a particular case is illustrated along with trajectories starting from the neighbourhood of the origin and spiralling to the cycle. The function used was $f(\dot{Y}) = \dot{Y} - \dot{Y}^3/3$, and the parameters were chosen so that $\lambda = 2$, $\kappa = 1$, $s = 0.5$, $v = 1.5$. Accordingly, (4) becomes

$$\ddot{Y} + Y = \dot{Y} - \dot{Y}^3. \quad (11)$$

The computation was obtained by a four point Runge-Kutta method. However, in the literature on perturbation methods referred to in the introduction, there are explicit calculation methods for successive approximations of a periodic solution once we know that it exists. Formally, the methods only work when the nonlinearity is small, but in practice quite good results are obtained with nonlinearities as large as in (11), or in any instance of the more general formulation (4). We only need more terms, as the relaxation oscillations occurring with large nonlinearities usually have important higher harmonics.

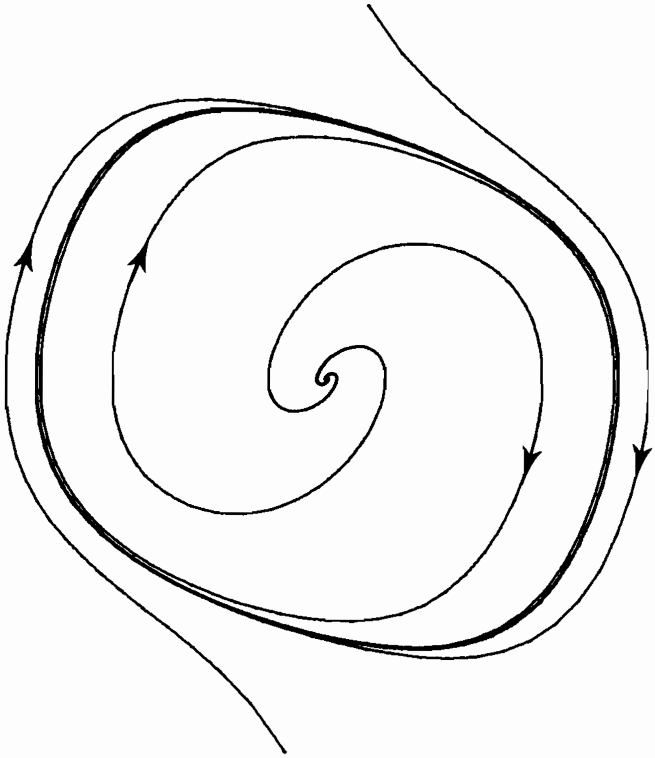


Figure 1   Limit cycle and some trajectories.

In an earlier publication, Puu (1986), the author used the two-timing method to obtain a first approximation, not only of the final limit cycle, but of the transient approach to it as well. Here we confine ourselves to the final limit cycle. The perturbation method used is the following. First we rewrite (11) so that the right hand side is multiplied by a "small" parameter ε. Thus

$$\ddot{Y} + Y = \varepsilon(\dot{Y} - \dot{Y}^3). \tag{12}$$

We try to find a solution in the form of a power series

$$Y(\tau) = Y_0(\tau) + \varepsilon Y_1(\tau) + \varepsilon^2 Y_2(\tau) + \ldots \tag{13a}$$

where

$$\tau = (1 + \varepsilon\omega_1 + \varepsilon^2\omega_2 + ...)t. \tag{13b}$$

We substitute (13a-b) in (12), so obtaining an equation expressed as a power series in ε equated to zero. We then equate all the coefficients to zero separately, obtaining a series of differential equations which can be solved in sequence. The first ones are

$$\ddot{Y}_0 + Y_0 = 0, \tag{14a-d}$$

$$\ddot{Y}_1 + Y_1 = -2\omega_1 \ddot{Y}_0 + \dot{Y}_0 - \dot{Y}_0^3,$$

$$\ddot{Y}_2 + Y_2 = -2\omega_1 \ddot{Y}_1 - (\omega_1^2 + 2\omega_2)\ddot{Y}_0 + \dot{Y}_1 + \omega_1 \dot{Y}_0 - 3\dot{Y}_0^2 \dot{Y}_1 - 3\omega_1 \dot{Y}_0^3,$$

and

$$\ddot{Y}_3 + Y_3 = -2\omega_1 \ddot{Y}_2 - (\omega_1^2 + 2\omega_2)\ddot{Y}_1 - 2(\omega_1\omega_2 + \omega_3)\ddot{Y}_0 + \dot{Y}_2 + \omega_1\dot{Y}_1 + \omega_2\dot{Y}_0$$

$$-3\omega_1^2 \dot{Y}_0^3 - 6\omega_1 \dot{Y}_0^2 \dot{Y}_1 - 3\dot{Y}_0 \dot{Y}_1^2 - 3\dot{Y}_0^2 \dot{Y}_2 - 3\omega_2\dot{Y}_0^3,$$

obtained for powers 0 through 3 of ε. Strictly, the series (13) converges for $\varepsilon \ll 1$, but quite good approximations can be obtained with ε as large as in our case (11). The calculations obviously become very complicated with ascending numbers of terms, so we will try to find a four-term approximation.

Formally, we have to settle the question of initial conditions. We are free to choose the origin of the time scale so that the rate of change of income is then zero. Accordingly

$$Y(0) = a_0 + \varepsilon a_1 + \varepsilon^2 a_2 + \varepsilon^3 a_3 + ... \tag{15}$$

and

$$\dot{Y}(0) = 0. \tag{16}$$

We note that we are not free to choose the initial amplitudes a priori. They must be chosen so that the composite periodic movement has the amplitude of the limit cycle. Approximating the limit cycle implies that we assume that we are on it already.

In the solution process, we always use the knowledge that the movement is periodic, and so we eliminate all terms which may result in non-periodic solutions. In order to leave sufficient freedom for adjustment of the basic period, the frequency is not fixed, but remains adjustable according to (13). Only the first unitary frequency has been fixed as it corresponds to the solution of the nonperturbed equation $\ddot{Y} + Y = 0$. Periodicity also implies that (15)-(16) may be written

$$Y_i(0) = a_i \tag{17}$$

$$\dot{Y}_i(0) = 0. \tag{18}$$

The first approximation (14a) is equal to the nonperturbed linear equation and so has a solution in terms of sin τ and cos τ. The initial conditions (17)-(18) settle it to

$$Y_0 = a_0 \cos \tau. \tag{19}$$

At the present stage a_0 is still underdetermined. From (19), we can readily calculate $\dot{Y}_0 = -a_0 \sin \tau$, $\ddot{Y}_0 = -a_0 \cos \tau$, and $\dot{Y}_0{}^3 = (3 \sin \tau - \sin 3\tau)/4$ and substitute in the right hand side of equation (14b). The latter then becomes:

$$\ddot{Y}_1 + Y_1 = 2\omega_1 a_0 \cos \tau - a_0 (1 - 3/4 a_0{}^2) \sin \tau - a_0{}^3/4 \sin 3\tau. \tag{20}$$

This linear equation can be solved easily. In the general solution the cos τ and sin τ terms would give rise to solution terms of the form τ cos τ and τ sin τ, which grow unboundedly and hence are nonperiodic. (Such terms first appeared in Astronomy and were called secular terms). As they violate the periodicity assumption, we have to make the coefficients of the cos τ and sin τ terms equal to zero. Discarding the trivial possibility that $a_0 = 0$, we obtain $\omega_1 = 0$ and $a_0 = 2/\sqrt{3}$.

By the periodicity assumption we have simultaneously estimated a_0 and ω_1, and this will hold in all the following steps of constructing the solution. We note that (19) becomes $Y_0 = 2/\sqrt{3} \cos \tau$, or

$$Y_0 = 1.1547 \cos \tau, \tag{21}$$

that (20) is simplified to

$$\ddot{Y}_1 + Y_1 = -2/\sqrt{27} \sin 3\tau, \tag{22}$$

and that $\tau = t$ in approximation number 1, ω_1 being zero.

The next step can be discussed more briefly. The general solution to (22), now without secular terms, is

$$Y_1 = a_1 \cos \tau + b_1 \sin \tau + 1/\sqrt{432} \sin 3\tau, \tag{23}$$

where from the initial conditions (16) we immediately get $b_1 = -1/\sqrt{48}$. Again, a_1 is still not determined until we have calculated the right hand side of (14c) and equated to zero the coefficients that would cause secular terms to appear in the solution for Y_2. At the same time ω_2 will be determined.

Carrying this out, we find $a_1 = 0$, $\omega_2 = -1/16$. Accordingly, (23) becomes

$$Y_1 = -0.1443 \sin \tau + 0.0481 \sin 3\tau. \tag{24}$$

We note that at this stage $\tau = t$, but at the next stage periodicity has to be corrected so that $\tau = 15/16\ t$.

We carry out two more steps:

$$Y_2 = 0.0812 \cos \tau + 0.0361 \cos 3\tau - 0.0060 \cos 5\tau. \tag{25}$$

$$Y_3 = 0.0135 \cos \tau + 0.0720 \sin \tau - 0.0139 \sin 3\tau - 0.0050 \sin 5\tau - 0.008 \sin 7\tau. \tag{26}$$

We also obtain that $\omega_3 = 0$, so we need not make any correction of the basic period, still having $\tau = 15/16\ t$. We find – comparing (21), (24), (25), and (26) – that the corrections of the amplitudes of the lower harmonics become smaller and smaller, as do the coefficients of the higher harmonics which are added at each stage. Accordingly, we can be confident that the series converges to the actual limit cycle, even though we have $\varepsilon = 1$. Of course, the convergence would have been much faster if the nonlinear perturbation had been really small. For $\varepsilon = 1$ our approximate solution is simply the sum $Y_0 + Y_1 + Y_2 + Y_3$.

In Figure 2 we illustrate the resulting approximated limit cycle in the phase plane, along with the limit cycle found by numerical methods and depicted in Figure 1.

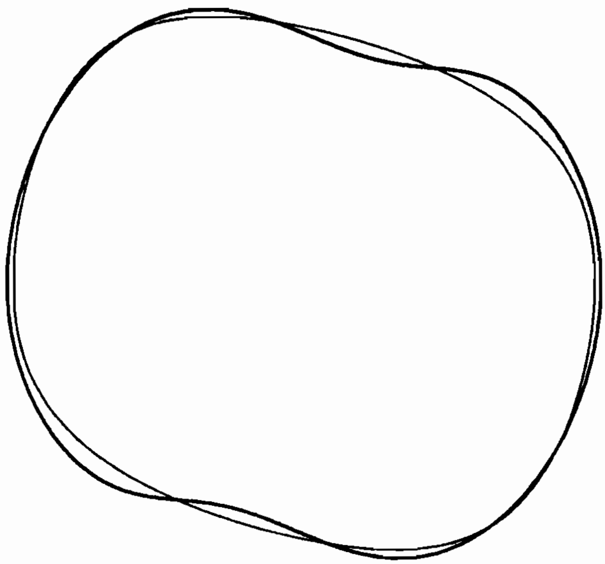


Figure 2   Limit cycle with four-term approximation

Details and general justifications of the perturbation method can be found in Kevorkian and Cole (1981). We should add that equations of the type (11) (called Rayleigh equations from Lord Rayleigh's early studies of bowed string instruments) are closely related to the van der Pol equations by mere differentiation and have thus been studied extensively in mathematics applied to electrical engineering. See Stoker (1950) and Hayashi (1964).

5. EXTENSION TO A TWO-REGION MODEL

In an earlier publication, Puu (1986), the author tried to show that really interesting phenomena occur when models of the business cycle are put in a spatial setting with interregional trade. This necessarily leads to a partial differential equation, which is not difficult to solve when the equation is linear. See Puu (1982), Beckmann and Puu (1985). However, the theory for nonlinear partial differential equations is far from well developed, so one needs to proceed with mere simulation, which necessarily becomes time consuming. An intermediate step would be to deal with two regions only. For the case of one big and one small country, where the influence is essentially unidirectional, we then deal with the case of a forced oscillator and have recourse to some interesting general studies.

Let us now consider what happens in an open economy. Then exports and imports must enter equation (1) and all the subsequent equations. Of course, we deal with two countries and a pair of coupled oscillators governed by equations like

$$\ddot{Y} + \lambda\kappa s_1 Y = \lambda\kappa v f(\dot{Y}) - (\kappa + \lambda s_1)\dot{Y} + \lambda\kappa m_2 Z - \lambda\kappa m_1 Y \tag{27}$$

$$\ddot{Z} + \lambda\kappa s_2 Z = \lambda\kappa v f(\dot{Z}) - (\kappa + \lambda s_2)\dot{Z} + \lambda\kappa m_1 Y - \lambda\kappa m_2 Z \tag{28}$$

where Z denotes the income of the second country. The indices on the propensities to save and to import refer to the countries. For simplicity, we can let all the other structural constants be the same in the two countries.

Now, suppose that the second country is very large compared to the first one.

Accordingly m_1 and m_2 are small and so is Y compared to Z. Thus, without any complications drop the m_1Y term in equation (28) which thus becomes autonomous, whereas (27) is complicated by a forcing term. Let us next suppose that the investment function is $v(\dot{Y} - \dot{Y}^3/3)$, and that the first approximation to the solution given by (19) is valid. Then in general

$$Z = A_2 \cos(\omega_2 t) + B_2 \sin(\omega_2 t) \tag{29}$$

where $\omega_2 = \sqrt{[\lambda\kappa(s_2 + m_2)]}$ and Y is governed by the differential equation

$$\ddot{Y} + \lambda\kappa(s_1 + m_1)Y = \lambda\kappa v(\dot{Y} - \dot{Y}^3/3) - (\kappa + \lambda s_1)\dot{Y} + \lambda\kappa m_2[A_2 \cos(\omega_2 t) + \\ + B_2 \sin(\omega_2 t)]. \tag{30}$$

This is still a Rayleigh equation, but now it contains a forcing term of frequency $\omega_2 = \sqrt{[(\lambda\kappa(s_2 + m_2)]}$, its own free frequency being $\omega_1 = \sqrt{[\lambda\kappa(s_1 + m_1)]}$. It is well known that in the case of an oscillator with a forcing term, there is nothing new if the forcing frequency is similar to the free frequency. Then the forced system still becomes a limit cycle. If the forcing frequency is substantially different, the phenomena of frequency entrainment, quasiperiodicity, and chaos may set in.

To simplify things, let $s_1 = 3/8$, $m_1 = 1/8$, $v = 3/2$, $\lambda = 4/3$, and $\kappa = 3/2$. Moreover, define $A = m_2A_2$, $B = m_2B_2$, and $\omega = \omega_2$. Then

$$\ddot{Y} + Y = \dot{Y} - \dot{Y}^3 + A \cos(\omega t) + B \sin(\omega t). \tag{31}$$

Obviously the natural frequency is unitary. It is easy to make projections for this system by using the same four-point Runge-Kutta method as before.

In Figure 3 we illustrate the range of possibilities as the driving frequency ranges from $\omega = 1.0$ to 3.6 with a step of 0.2, $\sqrt{(A^2 + B^2)}$ being equal to unity all the time. We can see how the system first produces limit cycles of the natural period of decreasing

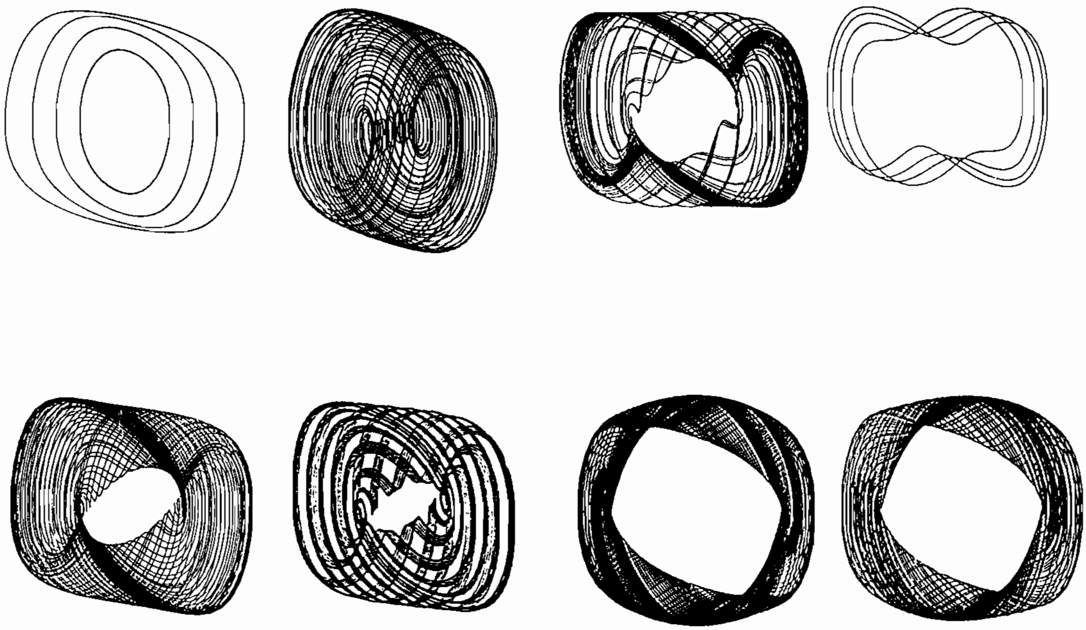


Figure 3 Behavior of forced nonlinear oscillator at various frequencies of the forcing term

amplitude as the driving frequency deviates more and more from the natural one. Finally, the system falls into a complex movement that covers a space of non-zero area measure. Still increasing the driving frequency, a hole opens up, until we finally get new limit cycles with harmonic entrainment around the driving frequency of 3. Further increasing the frequency, the system again falls into quasiperiodic or chaotic movement. We see that around the driving frequency of 2 there is decidedly no limit cycle, at least with the assumed unitary amplitude of the driving force.

Figure 4 shows what happens if we change the driving force amplitude from 1 to 1.2 and 1.5 respectively, keeping the frequency constant at 2. In the first case, a harmonic response seems to be superposed on the natural cycle; in the second, the natural frequency has been completely suppressed and the system has been entrained at the driving frequency. We conclude that variations in both frequency and amplitude of the periodic driving force can produce various kinds of complex combination cycles, as well as chaotic motion.

In the cases illustrated up to now we have dealt with driving frequencies equal to, or greater than unity. For integral frequencies, 2 and 3, we have seen that for suitable amplitudes there may be combination cycles of the natural and the driving frequencies. Another range of phenomena is revealed if we study the response to fractional driving frequencies, 1/2 and 1/3, illustrated in Figure 5 and 6. We find that entrainment can occur at these subharmonic frequencies as well. We can also see the very complicated limit cycles that remind us of the period doublings on the road to turbulence. See Feigenbaum (1983) and Abraham and Shaw (1982-85).

We shall now conclude this discussion of illustrative cases and proceed to discuss the matter in more systematic terms.

6. THE FORCED NONLINEAR OSCILLATOR

We shall use perturbation techniques to study the behaviour of (31), rewriting it thus:

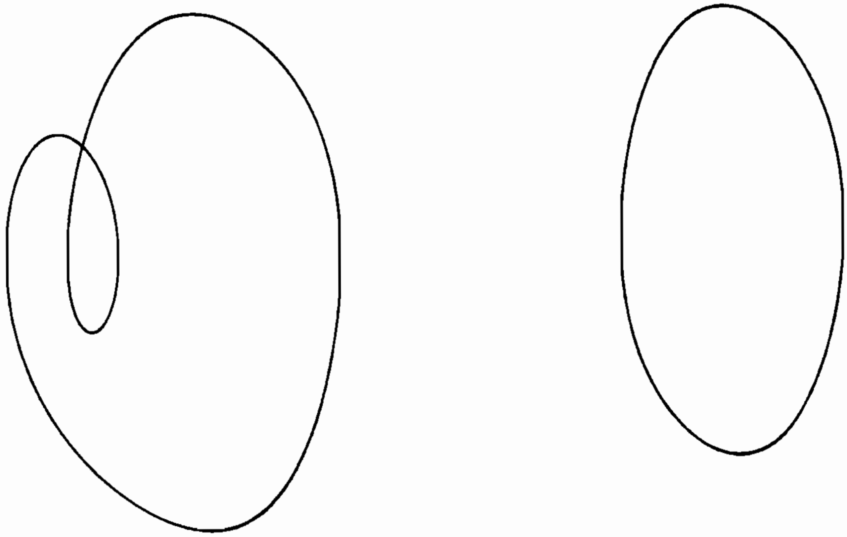


Figure 4   Various amplitudes (1.2 and 1.5) and frequency 2 of the forcing term

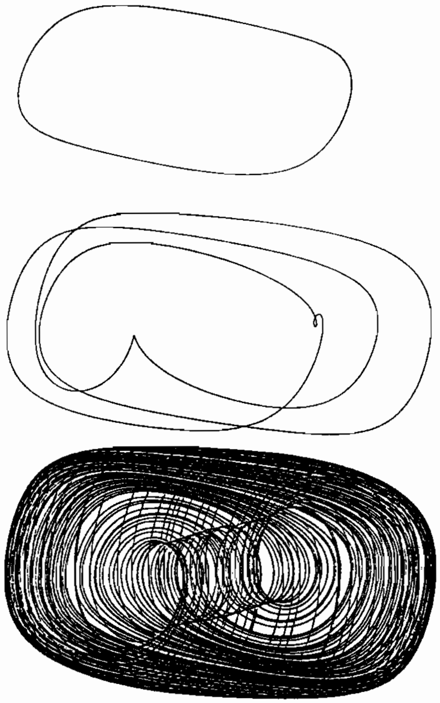


Figure 5   Various amplitudes (1.05/bottom/ to 1.25/top/ step 0.10) and driving frequency 1/2 of the forcing term

$$\ddot{Y} + Y = \varepsilon(\dot{Y} - \dot{Y}^3) + A\cos(\omega t) + B\sin(\omega t) \tag{32}$$

The size of ε is now a little more troublesome than it was before, because we are forced to work with the basic periodic solution alone, even though we know from (24)-(26) that the higher odd harmonics are by no means negligible. However, Hayashi (1964) has worked through the cases with large nonlinearities by means of an analog computer and he has been able to confirm that the results obtained for small nonlinearities still hold in a qualitative way, though not numerically, for the relaxation case.

For the case of harmonic response we try to find a solution to (32) of the form

$$Y = a(t)\cos\omega t + b(t)\sin\omega t \tag{33}$$

where the coefficients are now slowly varying functions of time. This assumption of slow variation is the basis of a set of approximations. Second derivatives $\ddot{a}$ and $\ddot{b}$ are deleted, as are all powers and products of the first derivatives $\dot{a}$ and $\dot{b}$. Likewise, products of the small quantities $\dot{a}$ and $\dot{b}$ with the "small" quantity ε are deleted. Finally as only the harmonic balance of the basic harmonic motion is studied, all terms involving $\cos 3\omega t$ and $\sin 3\omega t$ are left out. Under these assumptions we get

$$\varepsilon\dot{Y} = -\varepsilon\omega a\sin\omega t + \varepsilon\omega b\cos\omega t, \tag{34}$$

$$\varepsilon\dot{Y}^3 = 3/4\ \varepsilon\omega^3(a^2 + b^2)(b\cos\omega t - a\sin\omega t), \tag{35}$$

$$\ddot{Y} = -\omega^2(a\cos\omega t + b\sin\omega t) - 2\dot{a}\sin\omega t + 2\dot{b}\cos\omega t. \tag{36}$$

On substitution of (33)-(36) into (32), and equating the coefficients of $\cos\omega t$ and $\sin\omega t$ to zero separately, as we must in order that (32) be identically satisfied by our attempted solution, we obtain the two equations:

$$2\dot{a} = [1 - 3/4\ \omega^2\ (a^2 + b^2)]\ a + (1 - \omega^2)/\omega\ b - B/\omega, \tag{37}$$

$$2\dot{b} = -(1 - \omega^2)/\omega\ a + [1 - 3/4\ \omega^2\ (a^2 + b^2)]\ b + A/\omega. \tag{38}$$

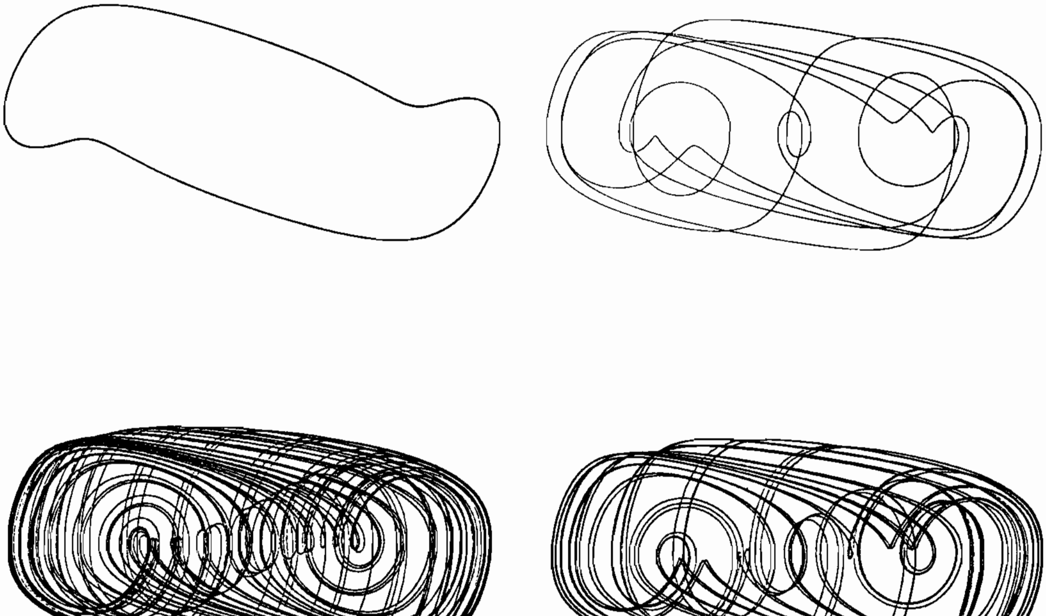


Figure 6 Various amplitudes (1.65/bottom/ to 1.80/top/ step 0.05) and driving frequency 1/3 of the forcing term

These are differential equations for the slowly varying amplitudes a and b. They are best studied by a series of transformations of variables and coefficients which are standard in the study of forced oscillators since the original work by van der Pol, Andronov and Witt. See Stoker (1950) or Hayashi (1964). Let $\alpha = \sqrt{3}\omega a/2$ and $\beta = \sqrt{3}\omega b/2$. We note that the new variables are ratios of a and b respectively to the natural amplitude of the equation without forcing terms, as found in (21). Moreover, they have been multiplied by the forcing frequency, which is not a common practice in the study of van der Pol's equation, but which simplifies matters greatly as we deal with an equation derived by differentiation of his own equation. The amplitudes of the forcing terms are likewise related to the amplitude of the natural basic oscillations by defining $F = -A\sqrt{3}/2$ and $G = -B\sqrt{3}/2$. Finally, we introduce the symbol σ for the "detuning",

$$\sigma = (1-\omega^2)/\omega \tag{39}$$

and define a new time scale $\tau = t/2$ to get rid of the multipliers of the left hand sides of (37)-(38).

We obtain

$$\dot{\alpha} = (1-\rho^2)\alpha + \sigma\beta + G \tag{40}$$

$$\dot{\beta} = -\sigma\alpha + (1-\rho^2)\beta - F \tag{41}$$

where

$$\rho^2 = \alpha^2 + \beta^2. \tag{42}$$

Let us keep in mind that ρ, unlike σ, is no constant, but the radius vector of a point in solution space for α, β. Accordingly, the differential equations (40)-(41) are nonlinear, containing cubic terms in the right hand sides.

According to van der Pol's original reasoning, harmonic entrainment according to the attempted solution (33) is correlated with equilibrium points for the differential equations (40)-(41), as determined by the algebraic equations

$$(1-\rho^2)\alpha + \sigma\beta = -G, \tag{43}$$

$$-\sigma\alpha + (1-\rho^2)\beta = F. \tag{44}$$

By squaring both sides of (43) and (44) and incorporating definition (42) we get the handy expression

$$[(1-\rho^2)^2 + \sigma^2]\,\rho^2 = F^2 + G^2. \tag{45}$$

Obviously, it determines the amplitude ρ for which α, β and, accordingly, the amplitudes a and b are in equilibrium. The solution to (45) may have from one to three real roots, depending on the detuning (dependent on frequency) and forcing amplitude constants. We will study the possibilities in more detail later after making some general comments.

In Figure 7 we show ρ^2 as a function of σ for various values of $F^2 + G^2$. These are the response curves to the forced oscillation. The figure also contains other curves which will be explained later. In the diagram we find the amplitude of the harmonically entrained oscillation, provided it can arise, i.e., if the corresponding equilibrium of (43)-(44) is stable.

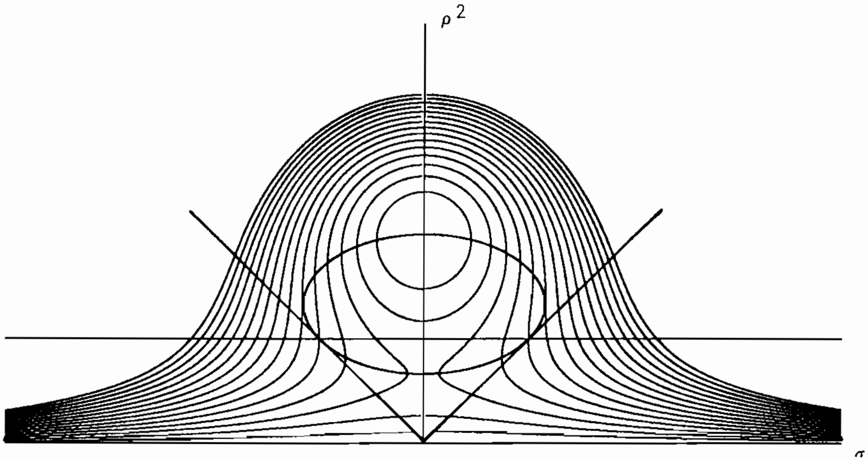


Figure 7 Response curves and stability region of the forced nonlinear oscillator

Harmonic entrainment is thus correlated with a stable equilibrium point for (40)-(41). We can now intuitively understand how a complex movement can arise.

Suppose that there is just one real root to (45) which is unstable. As the system (40)-(41) is obviously damped for large values of α and β, we can suspect that there is a limit cycle. Accordingly, the coefficients of the cyclic movement as assumed in (33) may themselves change cyclically. As there is no reason whatever why the frequency of the change of the coefficients should be in a rational relationship to ω, we may expect space-filling orbits to be quite a common occurrence.

We shall now study the stability of the possible stationary states of the system (40)-(41). For this reason, we linearize the system around a supposed equilibrium state, and define the new variables ξ and η as the deviations of α and β from their closest equilibrium values. In this way we get

$$\dot{\xi} = (1 - 3\alpha^2 - \beta^2)\xi - (2\alpha\beta - \sigma)\eta, \tag{46}$$

$$\dot{\eta} = -(2\alpha\beta + \sigma)\xi + (1 - \alpha^2 - 3\beta^2)\eta, \tag{47}$$

where, of course α and β are evaluated at the equilibrium point. The stability of equilibrium can now, according to Poincaré, be decided on the basis of the trace, the determinant, and the discriminant of these linear equations.

For the system (46)-(47) we obviously get

$$\text{Det} = (1 - \rho^2)(1 - 3\rho^2) + \sigma^2 \tag{48}$$

$$\text{Tr} = 2(1 - 2\rho^2) \tag{49}$$

and

$$\Delta = \text{Tr}^2 - 4\,\text{Det} = 4(\rho^2 + \sigma)(\rho^2 - \sigma). \tag{50}$$

The classification of equilibrium points is as follows. If Det < 0 then the point is a saddle point, having one stable and one unstable direction in the ξ,η- or α,β-plane. Consequently, the point is unstable. If Det > 0 then we have two cases, depending on the sign of the trace. When Tr < 0 we deal with a stable node or spiral; when Tr > 0 we deal with an unstable node or spiral. The sign of the discriminant just differentiates between nodes and spirals, depending on whether it is negative or positive. The most important conclusion is that for stability we must have Det > 0 and Tr< 0. See Jordan and Smith (1977).

The latter condition obviously implies that ρ^2 must be greater than 1/2. This border line between stability and instability is drawn as the horizontal straight line in Figure 7. Along with the straight line, there is an ellipse in the Figure, representing the locus of zeros for the determinant. Inside this ellipse, the determinant is negative, and we deal with saddles. Accordingly, the stable equilibrium points must be outside the ellipse and above the horizontal line. The diagram is due to Stoker (1950). Finally, there is a wedge representing the locus of points with $\rho^4 = \sigma^2$, i.e., points where the discriminant is zero. Above the wedge we deal with nodes (or saddles inside the ellipse), below it we deal with spirals.

The reader conversant with catastrophe theory may recognize that the pattern of response curves looks like a representation of a catastrophe surface of the cusp variety. For certain σ there are three values of ρ^2, the middle one always being inside the ellipse and hence unstable. So, if the "control parameters", σ and $F^2 + G^2$, are changed appropriately, we may experience sudden jumps of the "slow" amplitudes, irreversible processes, and other phenomena associated with the cusp catastrophe. See Gilmore (1981) or Poston and Stewart (1978).

There are also global characteristics which we may find for the non-linear equations (40)-(41). First, we note that for α, β sufficiently large, the third order terms will dominate, because σ and F, G are bounded. Obviously $\dot{\alpha}$ and $\dot{\beta}$ will have signs opposite to α and β respectively, and the system is damped. Accordingly, if we draw a closed curve sufficiently large to contain all the equilibrium points, then the sum of the indices of the equilibrium points must be unitary, as follows from the Poincaré Index Theorem. See Henle (1979) or Firby and Gardiner (1982). We recall that in our list of equilibrium points the index of a saddle is -1, that of a node or spiral 1 (whether stable or unstable).

However, from (43)-(44) and (45) we can see that the number of singularities is at most three, because (45) furnishes a third degree equation in ρ^2. It is helpful to calculate the discriminant D of this cubic equation. We get

$$27\,D = \sigma^2(1+\sigma^2)^2 + 27/4\,(F^2+G^2)^2 - (1-9\sigma^2)(F^2+G^2). \tag{51}$$

For $D > 0$ we have only one real root, for $D < 0$ we have three. With one equilibrium point, i.e. one root, we conclude that it must be a node or spiral according to the index theorem. With three roots, we have as the only possibility two nodes (or spirals) and one saddle.

Suppose we have just one singularity. We see that this case occurs with small driving forces, as then $D > 0$ according to (51). From (45) we see that it is then likely for ρ^2 to be small, which according to (49) means that the equilibrium point is an unstable node. However, as the system was damped at infinity it is likely that the amplitudes a and b in (33) themselves undergo a (slow) cyclic movement. When the two frequencies are incommensurable, then we get a spacefilling trajectory. That this occurs with small driving forces coincides with our findings from numerical solutions.

As another interesting possibility, suppose we have three equilibrium points, one stable node, one unstable node, and one saddle point. Then we may find that the trajectories of slowly changing amplitudes start at an unstable node, pass closely by the saddle, and finally end up at the stable node. As the a,b-system is passing the saddle it goes extremely slowly, being close to an equilibrium, and so the solution (33) may seem to settle down to a limit cycle with constant a and b, but finally it leaves the neighbourhood of the saddle point and starts settling down to the final stable equilibrium of (40)-(41), or limit cycle of (33).

A surprisingly rich set of possibilities can thus be found for the forced nonlinear oscillator, including quasiperiodicity, transitory limit cycles, and harmonic entrainment to the driving frequency. If we admit changes of the driving frequency and force, we can expect things like sudden collapses of limit cycles which have undergone continuous modifications only, and hysteresis.

We have limited the discussion to the case of harmonic entrainment and the phenomena around it. The discussion can also be extended to subharmonic entrainment. We shall not enter this matter. Reference should be made to Hayashi (1964) at this point.

7. COUPLED OSCILLATORS IN GENERAL

The case of the forced oscillator was arrived at by the simplifying assumption that the transmission of business cycles via trade was unidirectional, as in the case of a small open economy linked to the world market. However, it is certainly of interest to look at the more general case where no import propensities are zero. We can by no means be sure that even a weak coupling both ways does not alter the results substantially.

For simplicity of algebra let us assume that $\lambda\kappa v - \kappa - \lambda s_1 = \lambda\kappa v - \kappa - \lambda s_2 = 1$, and $\lambda\kappa v = 3$. Moreover, put $\lambda\kappa(s_1+m_1) = \omega_1^2$, $\lambda\kappa(s_2+m_2) = \omega_2^2$, and $\lambda\kappa m_1 = \mu_1$, $\lambda\kappa m_2 = \mu_2$.

Then (27)-(28) become

$$\ddot{Y} + \omega_1^2 Y = \dot{Y} - \dot{Y}^3 + \mu_1 Z \tag{52}$$

$$\ddot{Z} + \omega_2^2 Z = \dot{Z} - \dot{Z}^3 + \mu_2 Y \tag{53}$$

provided we again use the specification $f(x) = x - x^3/3$. The assumptions making the coefficients of the linear and cubic terms in the first derivatives unitary have no significance. They do not change the results from a qualitative point of view, and they simplify the algebra considerably. To let the accelerator terms be equal, of course, is a special case, but presently we wish to isolate the influences of the coupling amplitudes μ_i and the natural frequencies ω_i which were found to be crucial in the case of the forced oscillator.

In Figures 8 and 9 we show the results of numerical projection by a four-point Runge-Kutta method for the system (52)-(53). In both cases $\mu_1 = \mu_2 = 1$, whereas $\omega_1 = 1$, $\omega_2 = 2$ in the first case and $\omega_1 = 1$, $\omega_2 = 3$ in the second. We have arranged the phase portraits in the second and fourth quadrants in each picture, and placed a Lissajou figure, showing the coupling between the oscillators, in the first quadrant.

We shall now try to analyze the case with perturbation methods along the same lines that the forced oscillator was treated. So, we try to find a periodic solution with some unknown periodicity ω for both Y and Z, introducing slowly varying phases and amplitudes in terms of time dependent coefficients of the sine and cosine terms.

We try to find solutions of the form

$$Y = a_1(t) \cos \omega t + b_1(t) \sin \omega t, \tag{54}$$

$$Z = a_2(t) \cos \omega t + b_2(t) \sin \omega t, \tag{55}$$

and again use the approximations deleting all terms of second order and higher as in the case of (33) above. The equivalents of equations (37)-(38) are then

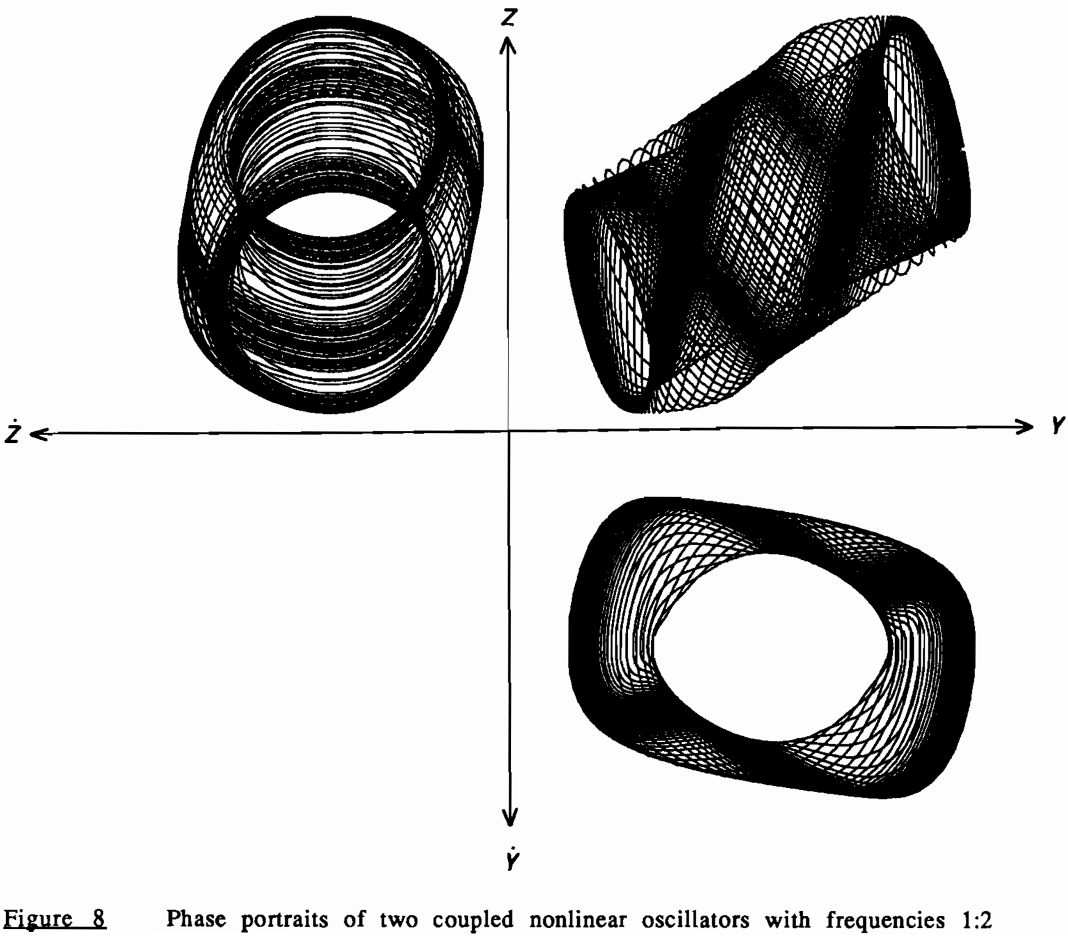


Figure 8 Phase portraits of two coupled nonlinear oscillators with frequencies 1:2

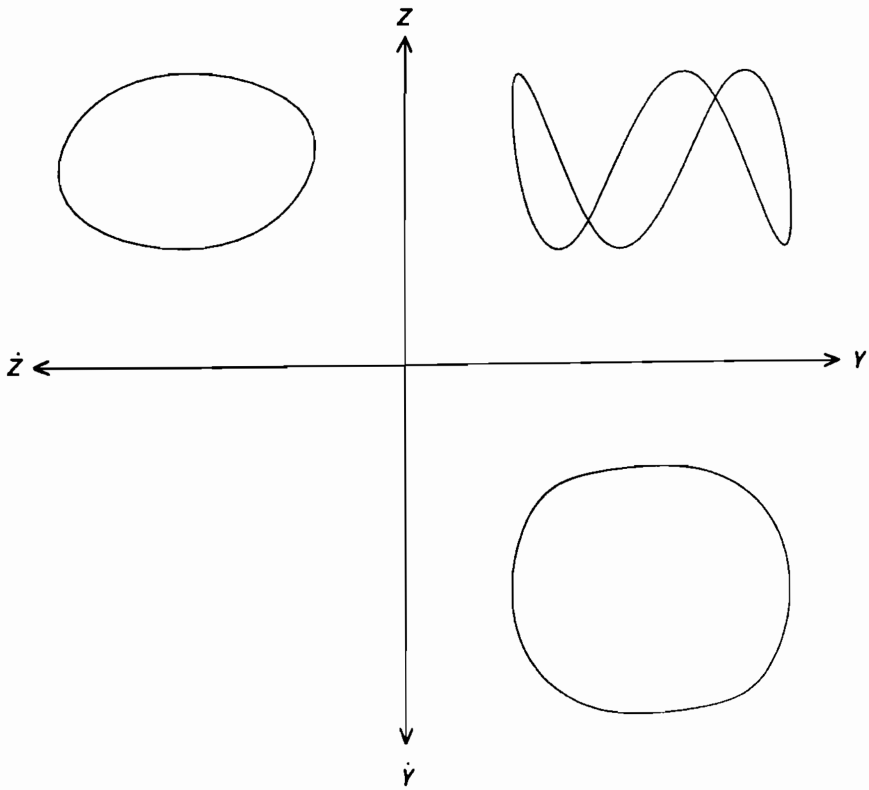


Figure 9 Phase portraits of two coupled nonlinear oscillators with frequencies 1:3.

$$2\,\dot{a}_1 = [1 - 3/4\ \omega^2(a_1{}^2 + b_1{}^2)]\,a_1 + (\omega_1{}^2 - \omega^2)/\omega\ \ b_1 - \mu_1\,b_2 \tag{56}$$

$$2\,\dot{b}_1 = -(\omega_1{}^2 - \omega^2)/\omega\,a_1 + [1\ -\ 3/4\ \omega^2(a_1{}^2 + b_1{}^2)]\,b_1 + \mu_1\,a_2 \tag{57}$$

$$2\,\dot{a}_2 = [1\ -\ 3/4\ \omega^2(a_2{}^2 + b_2{}^2)]\,a_2 + (\omega_2{}^2 - \omega^2)/\omega\,b_2 - \mu_2\,b_1 \tag{58}$$

$$2\,\dot{b}_2 = -(\omega_2{}^2 - \omega_2)/\omega\,a_2 + [1\ -\ 3/4\ \omega^2(a_2{}^2 + b_2{}^2)]\,b_2 + \mu_2\,a_1. \tag{59}$$

These expressions can also be simplified considerably by a change of notation. We define $\tau = t/2$, $\alpha_i = \sqrt{3}\,\omega a_i/2$, and $\beta_i = \sqrt{3}\,\omega b_i/2$. Moreover, let

$$\rho_i{}^2 = \alpha_i{}^2 + \beta_i{}^2, \tag{60}$$

and

$$\sigma_i = (\omega_i{}^2 - \omega^2)/\omega. \tag{61}$$

With this notation

$$\dot{\alpha}_1 = (1-\rho_1^2)\alpha_1 + \sigma_1\beta_1 - \mu_1\beta_2 \tag{62}$$

$$\dot{\beta}_1 = -\sigma_1\alpha_1 + (1-\rho_1^2)\beta_1 + \mu_1\alpha_2 \tag{63}$$

$$\dot{\alpha}_2 = (1-\rho_2^2)\alpha_2 + \sigma_2\beta_2 - \mu_2\beta_1 \tag{64}$$

$$\dot{\beta}_2 = -\sigma_2\alpha_2 + (1-\rho_2^2)\beta_2 + \mu_2\alpha_1. \tag{65}$$

Equations (62)-(63) and (64)-(65) are very similar to (40)-(41), except that the given forcing terms are replaced by linear coupling. The new equations are again of the third order, which means that for sufficiently large values of the variables the third order terms dominate. Accordingly, taking a sufficiently large hypersphere in the four-dimensional phase space of the system (62)-(65), we find that all trajectories point almost radially inwards. So the system is definitely damped at large values of the variables.

Inside such a hypersphere there may be up to six equilibrium points of the system, where the a_i and b_i in (54)-(55) are constant and there is a solution in terms of syncronized limit cycles for the variables Y and Z. Some of them may be unstable saddles or unstable nodes or spirals, and account for transitory limit cycles of Y and Z; some may be stable and lead to final limit cycles. Of course, we can also have a single unstable equilibrium which, in view of the damping at large amplitudes, would lead to a limit cycle for a_i, b_i. Again, if its frequency is incommensurable with ω, we can expect quasiperiodicity. Thus, the results conform with those we found by simulation. Synchronized cycles as well as drifting ones are possible for the coupled oscillators.

To push formal analysis a little further we equate (62)-(65) to zero, move the coupling terms to the other side of the equality signs, square and add the equations thus obtained pairwise. So,

$$[(1-\rho_1^2)^2 + \sigma_1^2]\,\rho_1^2 = \mu_1^2\rho_2^2 \tag{66}$$

$$[(1-\rho_2^2)^2 + \sigma_2^2]\,\rho_2^2 = \mu_2^2\rho_1^2 \tag{67}$$

corresponding to (45) are obtained. Substituting for one of the ρ_i^2 from the other obviously yields a sixth degree equation for the squared amplitude, which demonstrates the assertion above.

We can also see how this relates to our previous discussion of the forced oscillator case. Suppose $\mu_2 = 0$. Then (67) is independent of (66) and we find that - except for the trivial case where Z is at rest, i.e., $\rho_2 = 0$ - the only solution is $\sigma_2 = 0$, $\rho_2 = 1$. As there is no detuning we can see from (61) that $\omega = \omega_2$ and we find from the definitions of α_2, β_2 that $a_2^2 + b_2^2 = 4/3$, which corresponds to the natural amplitude. The right hand side of (66) is then constant and the analysis for this equation can follow the lines indicated for the forced oscillator.

Finally, we can study the stability of (62)-(65) around an equilibrium point by linearization . Defining ξ_i, η_i as the deviations from equilibrium of α_i, β_i we find

$$\dot{\xi}_1 = (1 - 3\alpha_1^2 - \beta_1^2)\ \xi_1 - (2\alpha_1\beta_1 - \sigma_1)\,\eta_1 - \mu_1\,\eta_2 \tag{68}$$

$$\dot{\eta}_1 = -(2\alpha_1\beta_1 + \sigma_1)\ \xi_1 + (1 - \alpha_1^2 - 3\beta_1^2)\,\eta_1 + \mu_1\,\xi_2 \tag{69}$$

$$\dot{\xi}_2 = (1 - 3\alpha_2^2 - \beta_2^2)\ \xi_2 - (2\alpha_2\beta_2 - \sigma_2)\,\eta_2 - \mu_2\,\eta_1 \tag{70}$$

$$\dot{\eta}_2 = -(2\alpha_2\beta_2 + \sigma_2)\ \xi_2 + (1 - \alpha_2^2 - 3\beta_2^2)\,\eta_2 + \mu_2\,\xi_1. \tag{71}$$

The stability around any equilibrium of the system (62)-(65) can now be studied by calculating the eigenvalues of the matrix of the linearized system (68)-(71). The use of stability criteria is straightforward, but a little messy, as we deal with a four by four matrix.

So far we have followed the same routes as in the case of the forced oscillator, where the forcing frequency was given a priori. According to (61) we get a pair of detuning coefficients σ_1, σ_2 for each choice of ω, and (66)-(67) then determine the amplitudes ρ_1 and ρ_2. It is likely that the investigations of stability would render whole intervals of ω and the other variables as possible solutions. Thus, we may have missed some additional piece of information.

As a matter of fact, the system (62)-(65) is not quite like (40)-(41) as the former is homogenous, but the latter is not. Inspection reveals that (62)-(65) has a zero solution, $\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 0$. To investigate its stability we could evaluate the eigenvalues of (68)-(71), but it is instructive to deal with the stability problem directly in terms of a suitable Lyapunov function.

To this end let us multiply (68)-(71) by the expressions $\mu_2\xi_1, -\mu_2\eta_1, -\mu_1\xi_2, \mu_1\eta_2$ respectively and add. Then we get

$$1/2\ d/dt\,[\mu_2(\xi_1^2 + \eta_1^2) - \mu_1(\xi_2^2 + \eta_2^2)] = \mu_2(\xi_1^2 + \eta_1^2) - \mu_1(\xi_2^2 + \eta_2^2). \quad (72)$$

This saddle dynamic for the amplitudes of the linearized system indicates that the origin is unstable. The amplitude which dominates initially will grow beyond any limit with the passage of time. Accordingly, we are not interested in the origin - i.e. a zero solution for (62)-65) - except when instability at the origin combined with damping for large amplitudes leads to a limit cycle.

How do we obtain a nonzero equilibrium for the system (62)-(65)? Due to a basic theorem in linear algebra, there is a condition for the system to possess such an equilibrium, namely that the matrix of (62)-(65) have a zero determinant. It does not matter that ρ_1 and ρ_2 are themselves dependent on the variables α_i and β_i. A necessary and sufficient condition for the existence of a nonzero stationary solution is still that the coefficient matrix, where $(1-\rho_1^2)$ and $(1-\rho_2^2)$ are treated as constants, has a zero determinant. Calculating this determinant and equating it to zero yields

$$(1-\rho_1^2)^2(1-\rho_2^2)^2 + \sigma_2^2(1-\rho_1^2)^2 + \sigma_1^2(1-\rho_2^2)^2 + (\mu_1\mu_2 - \sigma_1\sigma_2)^2 =$$
$$-2\mu_1\mu_2(1-\rho_1^2)(1-\rho_2^2), \quad (73)$$

which is a new expression independent of (66) and (67). It may thus be of some help for the determination of the resonance frequency ω.

Let us next multiply both sides of (66) and (67) and cancel the nonzero ρ_i^2. Multiplying out the parentheses we get

$$(1-\rho_1{}^2)^2(1-\rho_2{}^2)^2 + \sigma_2{}^2(1-\rho_1{}^2)^2 + \sigma_1{}^2(1-\rho_2{}^2)^2 + (\sigma_1\sigma_2)^2 = (\mu_1\mu_2)^2. \tag{74}$$

Subtracting (74) from (73)

$$(1-\rho_1{}^2)(1-\rho_2{}^2)/\sigma_1\sigma_2 = 1 - \mu_1\mu_2/\sigma_1\sigma_2 \tag{75}$$

is obtained. Finally, substitution from (75) into (74) yields

$$(1-\rho_1{}^2)^2/\sigma_1{}^2 + (1-\rho_2{}^2)^2/\sigma_2{}^2 = -2(1 - \mu_1\mu_2/\sigma_1\sigma_2). \tag{76}$$

These two expressions are useful. We can immediately see that substitution from (75) into (76) yields

$$(1-\rho_1{}^2)/\sigma_1 + (1-\rho_2{}^2)/\sigma_2 = 0. \tag{77}$$

Subtracting (75) twice from (76) we also get

$$(1-\rho_1{}^2)/\sigma_1 - (1-\rho_2{}^2)/\sigma_2 = 2\sqrt{}(\mu_1\mu_2/\sigma_1\sigma_2-1). \tag{78}$$

Equations (77)-(78) easily determine the amplitudes once the detuning coefficients are known. To determine these we have to consider that we have not used all the information contained in (66)-(67). From (77) we can see that the squares of the LHS terms are equal. Accordingly from (66) and (67) we have

$$\sigma_1{}^2\,\rho_1{}^4\,/\,\mu_1{}^2 = \sigma_2{}^2\,\rho_2{}^4\,/\,\mu_2{}^2,$$

or, in view of the fact that σ_1 and σ_2 have equal signs

$$\sigma_1\,\rho_1{}^2/\mu_1 - \sigma_2\,\rho_2{}^2/\mu_2 = 0. \tag{79}$$

It is now obvious that (77), (78), and (79) together with (61) determine the detuning constants, the amplitudes and the common frequency of the two coupled oscillators.

8. CONCLUSION

The more general spatial differential equation studied in Puu (1986), namely

$$\ddot{Y} + \lambda\kappa sY = \lambda\kappa v f(\dot{Y}) - (\kappa+\lambda s)\dot{Y} + \lambda\kappa m \nabla^2 Y,$$

where $\nabla^2 Y = \partial^2 Y/\partial x^2 + \partial^2 Y/\partial y^2$ is the Laplacian operator and x, y are the euclidean two-space coordinates, is much more difficult to analyse. The literature available on nonlinear partial differential equations is not encouraging. For a one-dimensional space, the case of $s = 0$ has been studied in connection with wind-induced oscillations in overhead lines. See Myerscough (1973, 1975). Preliminary simulations which the author has made with nonzero s indicate that the results obtained are of some relevance for our case. Anyhow, the case of continuous space is likely to introduce new interesting phenomena, some of them associated with the influence of boundary conditions.

Except for formal elegance, the author prefers continuous processes in time and space to discrete ones, due to the fact that discrete formulations are much more likely to produce abrupt and chaotic solutions. Accordingly, it is not very surprising to find them in discrete processes, and their occurrence may sometimes be associated with non-realistic features of the modelling in difference equations of low order.

As an example, let us consider the most elegant theory of population growth and diffusion due to Hotelling (1921). Only the section concerning the growth process concerns us here. If P denotes population, the process is formulated as $\dot{P} = K(S-P)P$, where K is some constant and S is the "saturation level" of population. The process is modelled after Malthusian principles where population first grows almost exponentially, due to natural multiplication. As the saturation size is approached, the means of living become more scarce and the process goes asymptotically to the saturation limit. The solution is the well known logistic curve and there is nothing mysterious about it.

On the other hand, the discrete counterpart of this process $P_t = K(S-P_{t-1})P_{t-1}$ is known as a prototype for chaos; see Feigenbaum (1983). Although the findings are most interesting, one may question the realism of this momentary mating for all individuals at completely regular intervals. It seems like the hog-cycle model in elementary economics, where all pig-owners make a decision at Christmas concerning the supply one year ahead. It is well known from oligopoly theory that any distribution over time of the responses of the individual agents tends to smooth out the most abrupt results of low order difference equations. In general, therefore, a continuous process more closely mimics reality.

The contention made here is that the complex dynamics described above is more interesting than the corresponding phenomena in discrete business cycle models. The latter can slip into the theoretical construction process due to bad modelling principles.

REFERENCES

Abraham, R.H. and Shaw, C.D., Dynamics: The Geometry of Behaviour I-IV , Aerial Press, Santa Cruz, 1982-84.

Allen, R.G.D., Mathematical Economics, Macmillan, New York, 1956.

Arnold, V.I., Geometrical Methods in the Theory of Ordinary Differential Equations, Springer, New York, 1983.

Beckmann, M.J. and Puu, T, Spatial Economics: Potential, Density, and Flow, North-Holland, Amsterdam, 1985.

Feigenbaum, M.J., 'Universal behaviour in nonlinear systems', in Barsenblatt, G.I., Iooss, G. and Joseph, D.D., (eds.), Nonlinear Dynamics and Turbulence, Pitman, London, 1983.

Firby, P.A. and Gardiner, C.F., Surface Topology, Ellis Horwood, Chichester, 1982.

Gilmore, R., Catastrophe Theory for Scientists and Engineers, Wiley, New York, 1981.

Goodwin, R.M., 'The nonlinear accelerator and the persistence of business cycles'. Econometrica 19, 1951, pp. 1-17.

Harrod, R.F., Towards a Dynamic Economics, Macmillan, London, 1948.

Hayashi, C., Nonlinear Oscillations in Physical Systems, Princeton University Press, Princeton, N.J., 1964.

Henle, M., A Combinatorial Introduction to Topology, Freeman, San Francisco, 1979.

Hicks, J.R., A Contribution to the Theory of the Trade Cycle, Oxford University Press, Oxford, 1950.

Hotelling, H.. 1921. A mathematical theory of migration, M.A. thesis presented at the University of Washington, republished 1978 in Environment and Planning A 10, 1978, pp. 1223-1239.

Jordan, D.W. and Smith, P.S., Nonlinear Ordinary Differential Equations, Oxford University Press, Oxford, 1977.

Kevorkian, J. and Cole, J.D., Perturbation Methods in Applied Mathematics, Springer, New York, 1981.

Myerscough, C.J., 'A simple model of the growth of wind-induced oscillations in over-head lines', Journal of Sound and Vibrations, 28, 1973, pp. 699-713.

Myerscough, C.J.,' Further studies of the growth of wind-induced oscillations in over-head lines', Journal of Sound and Vibration 39, 1975, pp. 503-517.

Poston, T. and Stewart, I., Catastrophe Theory and its Applications, Pitman, London, 1978.

Puu, T., 'Outline of a trade cycle model in continuous space and time', Geographical Analysis 14, 1982, pp. 1-9.

Puu, T., 'Multiplier-accelerator models revisited'; Regional Science and Urban Economics 16, 1986, pp. 81-95.

Stoker, J.J., Nonlinear Vibrations in Mechanical and Electrical Systems, Wiley, New York, 1950.

Chapter 11

Generalizations of Goodwin's Growth Cycle Model

J. GLOMBOWSKI and M. KRÜGER

1. INTRODUCTION

Goodwin's application of the predator-prey model from mathematical biology to theoretical economics has stimulated a considerable stream of follow-up work, including some of our own[1]. Of course, Goodwin was quite aware that his model was "starkly schematized and hence quite unrealistic" (Goodwin 1972, p. 442). While this statement provided ample opportunity for extensions and generalizations, the main stimulus for modifications should be based, at least in our view, on the basic cycle-generating mechanism of the model. In contrast to many earlier contributions by Goodwin to business cycle theory, this mechanism is not a Keynesian, but rather a Marxian one. This is because real wage increases are positively related to employment: whenever the rate of employment exceeds (falls short of) its 'normal' level, wage rises (declines) increase (decrease) the wage share in net product. On the other hand, accumulation and labour demand fall below (rise above) their 'normal' level, whenever the profit share does. This idea can be traced back to the 'general law of capitalist accumulation' (Marx 1974, pp. 574-606) which has also been discussed in more recent Marxist works (Bauer 1986, Sweezy 1942, Itoh 1980). However, it was not given a mathematical formulation until Goodwin combined the Marxian idea with Volterra's (1931) predator-prey model from mathematical biology.[2]

This employment-distribution mechanism can be seen as an important aspect of cyclical capital accumulation, but some other relevant aspects are missing. The model contains neither international economic relations, nor money and asset markets, and it ignores the economic role of the state, as do many real business cycle models. But it also neglects some typical Keynesian topics like product market disequilibria, and the divergence between real and money wages. The Marxian ideas of non-neutral technical progress and the social determination of labour productivity are also omitted.

This paper aims to analyse the economic effects of the latter aspects within a model of capital accumulation along Marxian lines. We shall consider the employment-distribution mechanism as the cycle-generating factor, and our modified system provides us with a kind of sensitivity test of the original model. Moreover, the modifications may make Goodwin's system more realistic. We shall focus upon four particular modifications:

(1) Production will not be determined by the size of the capital stock and a fixed capital-output-ratio, but rather by an adjustment mechanism with respect to excess demand.

(2) Work intensity may decline as the degree of employment rises, indicating an increase of workers' resistance in the labour process as their position vis à vis the capitalists is strengthened.

(3) Technical progress in the form of Harrod-neutrality is allowed to vary in speed, i.e., it shows simultaneous and identical changes in the growth rates of labour productivity and capital intensity. Additionally, technical progress may become non-neutral.

(4) Variations in the real wage rate result from assumptions concerning both money wage formation and price setting behaviour.

None of these modifications are new. The production adjustment mechanism has been used in Post-Keynesian business cycle theory, e.g. by Phillips (1954) and Bergstrom (1967). Variations in work intensity, induced by labour market conditions, have been emphasized by Radical economists (cf. Bowles, Gordon, Weisskopf 1983). Non-neutral technical progress occurs in Kaldor's growth theory as temporary deviations from the equilibrium point on the technical progress function (Kaldor 1957, p. 267), while permanent non-neutrality – in the form of a rising organic composition of capital – is a central theme in Marx (Marx 1974, p. 582ff.). Keynes emphasized the distinction between real wages and money wages in general, and the possibility of their inverse movement (Keynes 1973, p. 9ff.). Since then, the use of nominal wage functions, (such as the augmented Phillips-curve), together with price-setting assumptions, (e.g. the mark-up hypothesis), has become familiar. We shall proceed along the lines suggested by Solow and Stiglitz (1968).

The structure of the paper is as follows: In section 2 the general version of the model, employing all modifications simultaneously, is introduced. Section 3 is devoted to a simplified case which gives rise to a Goodwin-type solution, i.e., a closed solution curve of the wage share in income and the rate of employment in the phase space. While the assumption of neutral technical progress is preserved, the effects of a varying speed of technical progress and of variations in work intensity are considered in section 4. Section 5 deals with non-neutral technical progess under simplified assumptions, while the general model is taken up again in section 6. Finally, the results are summarized and further possible modifications are suggested.

2. THE GENERAL MODEL

In formulating the general model, we have to introduce a large set of equations, variables and parameters. Although this may appear cumbersome, it seems to be unavoidable in order to clarify our main assumptions. All equations will be given en bloc first and, subsequently, will be discussed in some detail. While the symbols are defined immediately after the equations, for convenience the reader will also find a list of symbols in the Appendix. The model is formulated in continuous time. Time derivatives are denoted by a dot, e.g. $dx/dt \equiv \dot{x}$, while growth rates will be indicated by a hat, e.g. $\dot{x}/x \equiv \hat{x}$.

The general version of the model consists of the following nineteen equations:

$$P = \sigma K \tag{2.1}$$

$$\Gamma = P/\gamma \tag{2.2}$$

$$Y = \theta P \tag{2.3}$$

$$L = h\theta\Gamma \tag{2.4}$$

$$v = K/Y \tag{2.5}$$

$$y = Y/L \tag{2.6}$$

$$(P/\hat{\ }\Gamma) = \mu_1 + \mu_2 (K/\hat{\ }\Gamma) \tag{2.7}$$

$$\omega = w/p \tag{2.8}$$

$$\lambda = \omega/y \tag{2.9}$$

$$(\hat{K/\Gamma}) = \nu_1 + \nu_2 \lambda \tag{2.10}$$

$$\beta = L/A \tag{2.11}$$

$$\hat{A} = n \tag{2.12}$$

$$\hat{h} = -\gamma_1 + \gamma_2 \beta = \varphi(\beta) \tag{2.13}$$

$$\hat{w} = -a_1 + a_2 \beta + a_3 \hat{p} \tag{2.14}$$

$$\hat{p} = -b_1 + b_2 \theta + b_3 \hat{w} \tag{2.15}$$

$$\Pi = (1-\lambda) Y \tag{2.16}$$

$$\dot{K} = c \Pi \tag{2.17}$$

$$S = s_w \lambda Y + s_\pi \Pi \tag{2.18}$$

$$\dot{Y} = \delta(K - S) \tag{2.19}$$

Equation (2.1) postulates a technical relation between the capital stock (K) and maximum net production (P) obtainable it in the absence of a labour shortage. We might call σ the 'technical capital productivity'. As we shall see later, it is assumed to be a variable. (2.2) is a second technical assumption. It relates maximum net production to labour requirements (Γ), necessary to obtain maximum production under conditions of maximum work intensity. We might call γ the 'technical labour productivity'. It will also be considered as a variable. Our model allows for changes in capacity utilization and work intensity. Therefore, 'technical' and 'actual' factor productivities will generally differ. Equation (2.3) defines the rate of capacity utilization (θ) as the share of actual net output (Y) in maximum net output. If capacity utilization falls short of its maximum, only a part of the available capital stock is used and - under constant returns to scale – only $\theta\Gamma$ units of labour are required to produce Y under conditions of maximum work intensity. Now let h be a labour inefficiency index which is equal to 1 if work intensity is at its maximum, and rises in proportion to decreases in work intensity. As equation (2.4) shows, the actual necessary labour input or employment (L), is linked to labour input under full capacity utilization and maximum work intensity by the rate of capacity utilization and the labour inefficiency index. Equations (2.5) and (2.6) describe the standard definitions of the (actual) capital coefficient (v) and (actual) labour productivity (y). Figure 1 provides a schematic account of the relations between the variables introduced so far.

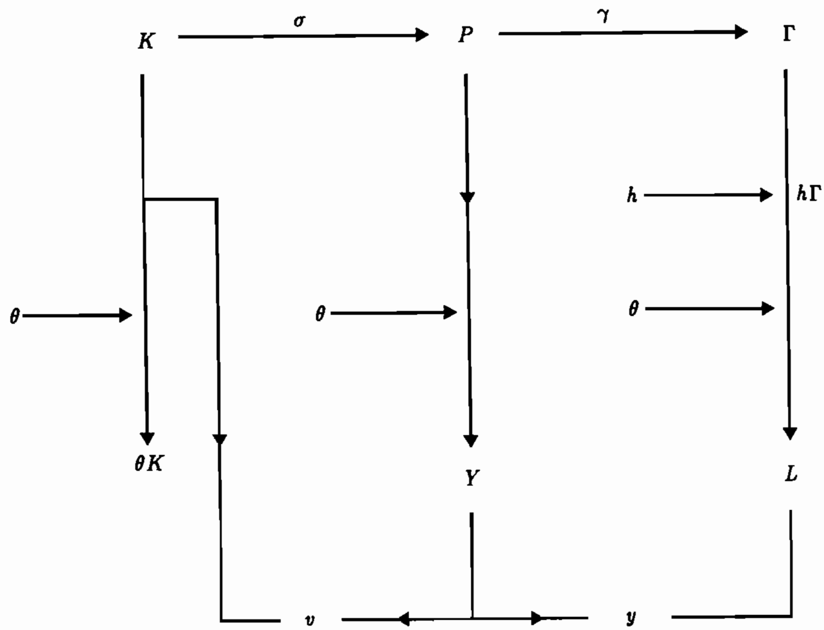


Figure 1 Relations between variables involved in the technical progress function

For the sake of clarity, let us state the relations between the 'technical' and the 'actual' versions of the capital coefficient, labour productivity, and capital intensity. Taking the definitions into account it follows that:

$$v = \frac{K}{Y} = \frac{K}{P}\,\frac{1}{\theta} \tag{2.20}$$

$$y = \frac{Y}{L} = \frac{P}{\Gamma}\,\frac{1}{h} \tag{2.21}$$

$$\frac{K}{L} = \frac{K}{\Gamma}\,\frac{1}{h\theta} \tag{2.22}$$

These distinctions are quite useful. Consider (2.7), which is a modified version of Kaldor's technical progress function in its linear form: the growth rate of technical capital productivity is assumed to depend linearly on the growth rate of technical capital intensity. Kaldor defined his original function in terms of the growth rates of the actual variables, i.e., $(\hat{Y/L})$ and $(\hat{K/L})$. Now, applying (2.21), (2.22), and (2.7), we can show that our assumptions imply, in Kaldor's variables, the technical progress function

$$\hat{Y/L} = \mu_1 + \mu_2 \hat{K/L} - (1 - \mu_2)\hat{h} + \mu_2 \hat{\theta}. \tag{2.23}$$

Therefore Kaldor's function does not seem to be well-defined in accumulation models which include variable capacity utilization and work intensity. Unless these variables are kept constant, his function is subject to permanent shifts. Now these problems can be avoided if the economic hypothesis underlying the technical progress function is expressed in terms of technical variables as in (2.7).

Equation (2.8) defines the real wage rate (ω) as the fraction of the money wage rate (w) and the price level (p). The wage share in net product (λ) is then defined as the fraction of the real wage rate and actual labour productivity, as in (2.9). A 'mechanization function' is introduced in (2.10). According to this function, the growth rate of technical capital intensity is related to income distribution in the sense that mechanization is encouraged by a high wage share, i.e., high labour costs per unit of net product.

The technical progress function and the mechanization function are both depicted in Figure 2. According to the assumptions about the four parameters μ_1, μ_2, ν_1 and ν_2, one can distinguish various cases. Consider for instance:

(a) $\mu_1 = 0$, $\mu_2 = 1$.

In this case technical capital productivity σ will remain constant. However, this assumption leaves room for different speeds of technical progress because the growth rates of technical labour productivity and technical capital intensity, which are equal, may become higher or lower, depending on the level of the wage share. Technical progress would occur at a constant rate only if, in addition, $\nu_2 = 0$.

Another interesting case is

(b) $\mu_1 = 0$, $0 < \mu_2 < 1$.

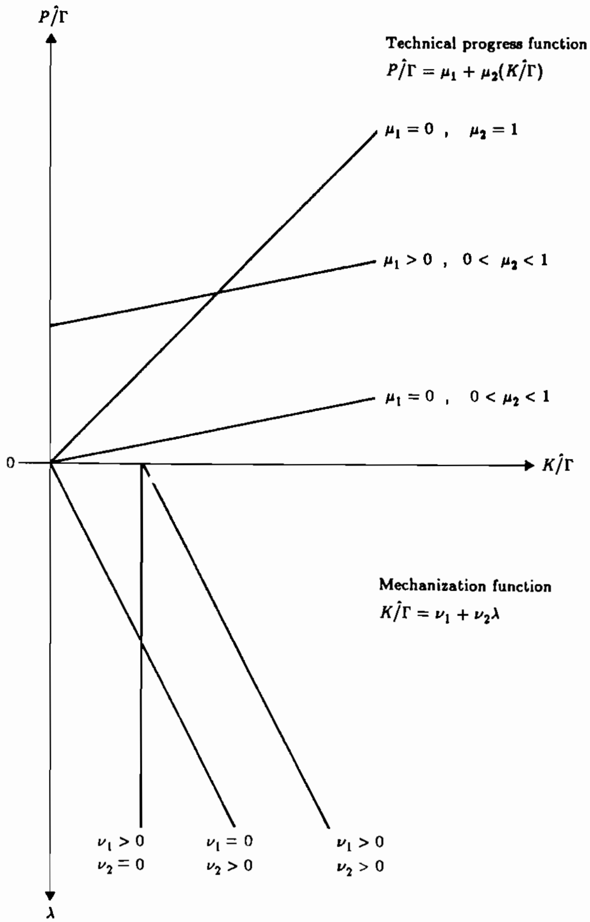


Figure 2 The technical progress function and the mechanization function

In this situation technical capital productivity (σ) is bound to fall steadily. An exception is the special case $\nu_1 = 0$ in combination with a vanishing wage share. Type (b) as a technical progress function could be labelled 'Marxian', because it reproduces the Marxian assumption of the permanent rise of the technical composition of capital.

Finally, we may consider the specification

(c) $\mu_1 > 0$, $0 < \mu_2 < 1$,

which can be called 'Kaldorian' as it provides for the possibility of a constant technical productivity in equilibrium. An appropriate positive steady-state wage share would exist, while fluctuations of income distribution around its equilibrium value would give rise to phases of non-neutral technical progress of either capital saving or labour saving type.

The rate of employment (β) is defined in (2.11). It is the fraction of actual labour requirements (employment) to labour supply (A). As may readily be seen, we apply in (2.12) the standard assumption of constant growth (with the rate n) for the supply side of labour. The rate of employment affects work intensity in a way described by (2.13). The labour inefficiency index will rise whenever the rate of employment is 'high' ($\beta > \gamma_1/\gamma_2$) and h will decline if $\beta < \gamma_1/\gamma_2$. This assumption tries to reflect the influence of the employment situation on workers' capability to resist work pressure by various means of restrictive practices. As the growth rate of the labour inefficiency index increases with rising employment, the rise of actual productivity would lag behind that of technical labour productivity in a boom. This negative effect on productivity could be compensated for by a reduction of labour hoarding in the course of a boom. However, in the present model we have decided to neglect this significant effect.[3)]

A rising rate of employment is also assumed to affect wage increases. While Goodwin stipulated a real wage growth/employment relation, equation (2.14) indicates that the growth rate of nominal wages ($\hat{w}$) depends on the rate of employment. Moreover, the inflation rate ($\hat{p}$) is assumed to contribute to wage increases too, because workers try (and succeed) to get partial compensation for price increases. On the other hand, the inflation rate is thought to depend on wage increases as well as on capacity utilization, cf. (2.15). It is assumed that capitalists react to rising costs and rising demand by raising the prices of their products.

Profits (Π) are defined as residual, i.e., non-wage income; see (2.16). Therefore, we implicitly define income such that it is equal to the net product.[4)] In addition we assume that a constant share (c) of profits is invested, i.e., transformed into capital ($\dot{K}$), cf. (2.17). Savings (S) are described by Kaldor's saving function, stated as (2.18), where s_w and s_π, $0 < s_w < s_\pi < 1$, denote the specific saving ratios out of wages and profits, respectively. Savings and investment plans are uncoordinated, which means that their planned volumes may differ. The realized values may also differ because we assume that planned investment can be realized, if necessary, by making use of the stock of investment goods. As our definition of the capital stock does not include unplanned inventories, the standard ex post identity between savings and investment does not apply. Differences between savings and investment are equivalent to differences between net production (=income) and demand. Equation (2.19) describes a rather standard adjustment principle. Production will be adjusted to either excess demand ($\dot{K} > S$) or to excess supply ($\dot{K} < S$). The constant parameter δ may be called the 'production adjustment coefficient'.

It will shed some light on the character of our model and its variation from the original Goodwin model to specify the parameter restrictions under which our model would be transformed into its original form:

(1) For $c = s_\pi = 1$ and $s_w = 0$, Goodwin's identity of profits, savings and investment would be restored. The production adjustment equation (2.10) would have to be omitted, and capacity utilization (θ) should also be equal to one in order to determine the actual production level (Y) by its maximum (P).

(2) Work intensity must be kept constant. Therefore both parameters γ_1 and γ_2 have to be equal to zero, cf. (2.13); moreover, the labour inefficiency index must be fixed at a certain level, e.g. $h = 1$.

(3) Technical progress must become Harrod-neutral and proceed with a constant rate of growth. Thus $\nu_1 > 0, \nu_2 = 0$ (cf. (2.10)) and $\mu_1 = 0, \mu_2 = 1$ (cf. (2.7)). In other words, the technical capital intensity would grow with an exogenously given rate and the technical progress function would be identical with the 45°-line.

(4) Finally, the restriction $b_1 = b_2 = b_3 = 0$ would provide for a constant price level; cf. equation (2.15). The rate of employment (β) would once again determine real wage rather than money wage changes.

Our general parameter restrictions given below can be seen to allow for a revival of Goodwin's assumptions with one exception: The production adjustment mechanism - and thus later the variability of capacity utilization - is preserved. The other possibilities for simplifications will be used in various combinations in the later sections. Throughout the paper, unless stronger restrictions are explicitly imposed, we apply the following set of parameter restrictions:

$$\mu_1 \geqq 0\ ,\ 1 \geqq \mu_2 > 0 \tag{2.24}$$

$$\nu_1 \geqq 0\ ,\ \nu_2 > 0 \tag{2.25}$$

$$n \geqq 0 \tag{2.26}$$

$$\gamma_1 \geqq 0\ ,\ \gamma_2 \geqq 0 \tag{2.27}$$

$$a_1 > 0\ ,\quad a_2 > 0\ ,\quad 1 \geqq a_3 \geqq 0 \tag{2.28}$$

$$b_1 \geqq 0\ ,\quad b_2 \geqq 0\ ,\quad 1 \geqq b_3 \geqq 0 \tag{2.29}$$

$$1 > c > 0 \tag{2.30}$$

$$1 > s_\pi > s_w > 0 \tag{2.31}$$

$$\delta > 0 \tag{2.32}$$

$$1 > a_3 b_3 \tag{2.33}$$

$$g: = c - (s_\pi - s_w) > 0\ . \tag{2.34}$$

Only the last two restrictions need some explanation. We assume (2.33) to preclude the wage – price spiral running into hyperinflation. By (2.34) it is assumed that a rise in the profit share (Π/Y) will induce an increase of the investment share ahead of the savings ratio. From (2.16) to (2.18) we get

$$\dot{K}/Y - S/Y = (c-(s_\pi - s_w))\Pi/Y \tag{2.35}$$

which in turn implies

$$\frac{d\,(\dot{K}/Y - S/Y)}{d\,(\Pi/Y)} > 0 \tag{2.36}$$

cf. (2.34).

It should be emphasized that "underconsumptionist" thinking would rather postulate the opposite; i.e. that a rising wage share would stimulate demand and expansion of production. Our present assumption seems to fit 'profit squeeze' versions of Marxist cycle theory more closely. [5)]

We can now take two steps to reduce the model to a system of only four equations. The first step consists of the transformation of the original functions into equations containing growth rates. From (2.6), (2.11), and (2.12) we obtain

$$\hat{\beta} = \hat{Y} - y - n \tag{2.37}$$

while (2.9) implies

$$\hat{\lambda} = \hat{\omega} - y. \tag{2.38}$$

The growth rate of actual production ($\hat{Y}$) is found from (2.16)-(2.19) to be a function f of the wage share:

$$\hat{Y} = \delta[g(1-\lambda) - s_w] = f(\lambda). \tag{2.39}$$

Note that, as a consequence of (2.34), the growth rate of production declines as the wage share rises.

Next, we shall express the growth rate of actual labour productivity by means of the variables affecting it. From (2.21) we get

$$\hat{y} = (\widehat{P/\Gamma}) - \hat{h} \tag{2.40}$$

This equation demonstrates that the growth rate of technical labour productivity has to be corrected by changes in work intensity to obtain actual productivity. Now $\hat{h}$ is a function of the rate of employment, and $\widehat{P/\Gamma}$ can be shown – from the technical progress function and the mechanization function – to be a function u of the wage share

$$\widehat{P/\Gamma} = u(\lambda) = \mu_1 + \mu_2\nu_1 + \mu_2\nu_2\lambda. \tag{2.41}$$

Consider the growth rate of the real wage rate which appears in (2.38). It follows from its definition, as well as from the price and money wage functions, (2.14) and (2.15), that real wage growth depends on the rate of employment (β) and capacity utilization (θ). We find that

$$\hat{\omega} = \Phi_1(\beta) - \Phi_2(\theta) \tag{2.42}$$

where $\Phi_1(\beta)$ and $\Phi_2(\theta)$ are the following simple expressions:

$$\Phi_1(\beta) = [b_1(1 - a_3) - a_1(1 - b_3) + a_2(1 - b_3)\beta\ ] / (1 - a_3 b_3) \tag{2.43}$$

and

$$\Phi_2(\theta) = b_2(1 - a_3)\theta/(1 - a_3 b_3). \tag{2.44}$$

Unfortunately, capacity utilization cannot be expressed directly in terms of the wage share and/or the rate of employment. Its growth rate, however, follows from (2.20) as the difference of the growth rates of the technical and the actual capital coefficients:

$$\hat{\theta} = \widehat{K/P} - \hat{v}. \tag{2.45}$$

From the technical progress function and the mechanization function, K/P is found to be a function ψ of λ:

$$\widehat{K/P} = \psi(\lambda) = \nu_1 - \mu_1 - \mu_2\nu_1 + \nu_2(1 - \mu_2)\lambda. \tag{2.46}$$

(2.5) implies

$$\hat{v} = \hat{K} - \hat{Y}. \tag{2.47}$$

Finally, the growth rate of the capital stock is derived from (2.16), (2.17), and (2.5) as a function z of the wage share and the actual capital coefficient

$$\hat{K} = z(\lambda, v) = c(1 - \lambda)/v. \tag{2.48}$$

We are now prepared to show that the model can be reduced to a system of four non-linear differential equations in the rate of employment, the wage share, the rate of capacity utilization, and the actual capital coefficient. We state the equations en bloc first and comment afterwards on their derivation.

$$\dot{\beta} = [\varphi(\beta) + f(\lambda) - u(\lambda) - n]\,\beta \tag{2.49}$$

$$\dot{\lambda} = [\varphi(\beta) + \phi_1(\beta) - u(\lambda) - \phi_2(\theta)]\,\lambda \tag{2.50}$$

$$\dot{\theta} = [\psi(\lambda) + f(\lambda) - z(\lambda,v)]\,\theta \tag{2.51}$$

$$\dot{v} = [-f(\lambda) + z(\lambda,v)]\,v \tag{2.52}$$

The differential equation for the rate of employment (2.49) follows from (2.37) by successive substitutions from (2.39) - (2.41), and (2.13). Likewise, we obtain the equation for the wage share (2.50) by replacing the elements of the right-hand side of (2.38) by (2.40) - (2.44) and (2.13). As far as equation (2.51) is concerned, we start from (2.45) and take into account (2.46) - (2.48) as well as (2.39). Finally, the equation for the actual capital coefficient (2.52) follows from (2.47), (2.48), and (2.39).

In the general case, i.e. operating with all modifications, the model cannot be reduced further. This is due to the fact that capacity utilization influences price formation and, by virtue of that, enters equation (2.50) as one of the variables governing the movement of the wage share. Otherwise, (2.49) and (2.50) would form an independent subsystem in the rate of employment and the wage share only.

Before considering the model in its most general version (see section 6), some simplified cases shall be discussed in the next three sections.

3. A SIMPLIFIED GOODWIN-LIKE CASE

Under various simplified assumptions a result which is rather similar to that of Goodwin's original model can be established. In this section we ignore the effect of the rate of employment on work intensity:

$$\varphi(\beta) = 0 \tag{3.1}$$

Secondly, by assuming that

$$\mu_1 = 0\ ,\ \ \mu_2 = 1\ ,\ \ \nu_1 > 0\ ,\ \ \nu_2 = 0 \tag{3.2}$$

we exclude all forms of technical progress except the neutral one at constant pace.

This implies that

$$\psi(\lambda) = \hat{K/P} = 0 \quad \text{and} \tag{3.3}$$

$$u(\lambda) = \hat{P/\Gamma} = \nu_1 . \tag{3.4}$$

Finally, we remove any influence of capacity utilization on distribution; i.e. we stipulate that

$$\phi_2(\theta) = 0. \tag{3.5}$$

(3.5) can be assured by assuming, alternatively,

$$b_2 = 0 \tag{3.6}$$

or

$$a_3 = 1. \tag{3.7}$$

The latter assumption means that workers are able to defend themselves against price increases by obtaining wage rises instantaneously at the same rate. The former assumption implies that capacity utilization has no influence on price formation. While both alternatives are restrictive, the reader may choose for himself the less unrealistic one.[6] We shall proceed by applying (3.6). Note that, if (3.7) was assumed in addition, some of the formulae would become much simpler.

As may readily be checked, the system (2.40) - (2.52) is now transformed into

$$\dot{\beta} = [f(\lambda) - \nu_1 - n]\, \beta \tag{3.8}$$

$$\dot{\lambda} = [\phi_1(\beta) - \nu_1]\, \lambda \tag{3.9}$$

$$\dot{v} = [-f(\lambda) + z(\lambda, v)]\, v \tag{3.10}$$

$$\theta = 1/(\sigma v) \quad , \quad \sigma = \text{const.} \tag{3.11}$$

While equations (3.8) – (3.10) need no explanation, we should comment briefly on (3.11). It is simply a consequence of equation (3.3). If the technical capital coefficient is assumed to be constant, its inverse, the technical capital productivity σ, must also be constant. Then, from (2.20) we get capacity utilization as a simple function of the actual capital coefficient.

The system (3.8) – (3.11) has a steady state solution. Moreover, it is decomposable because (3.8) and (3.9) depend on β and λ only. The equilibrium values are as follows:

$$\lambda_e = [\delta(g - s_w) - (\nu_1 + n)] / (\delta g) \tag{3.12}$$

$$\beta_e = [\nu_1(1 - a_3 b_3) + a_1(1 - b_3) - b_1(1 - a_3)] / [a_2(1 - b_3)] \tag{3.13}$$

$$v_e = [c(1 - \lambda_e)] / [\delta g(1 - \lambda_e) - s_w] \tag{3.14}$$

$$\theta_e = 1/(\sigma v_e). \tag{3.15}$$

As to the dynamic behaviour of the present case, consider the Jacobian of the independent subsystem (3.8) – (3.9), linearized around the equilibrium point, i.e.,

$$\begin{bmatrix} 0 & B \\ C & 0 \end{bmatrix}$$

where

$$B = f'(\lambda_e)\, \beta_e \tag{3.16}$$

$$C = \phi_1'(\beta_e)\, \lambda_e. \tag{3.17}$$

This gives rise to a pair of purely imaginary eigenvalues

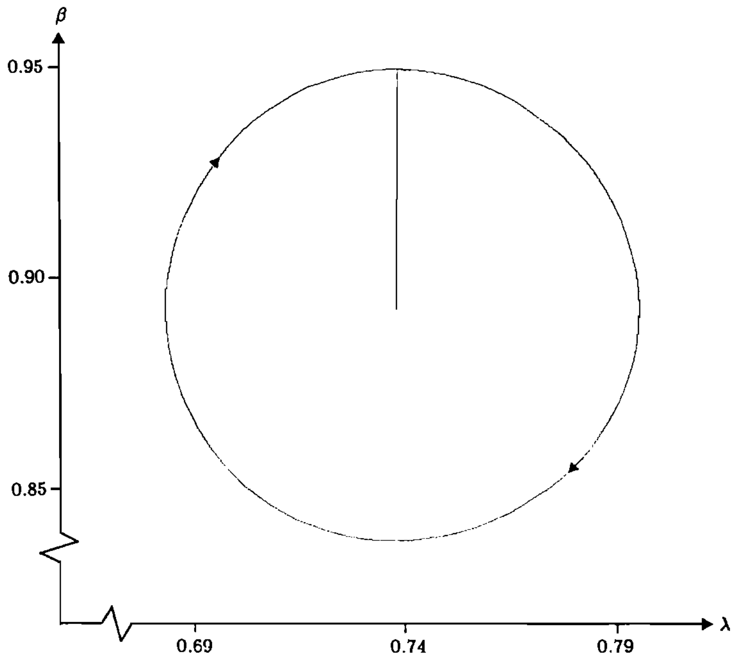
$$s_{1,2} = \pm\ (BC)^{1/2} = \pm\ i\ [(\delta g a_2 (1 - b_3)\ \beta_e \lambda_e) / (1 - a_3 b_3)]^{1/2}. \tag{3.18}$$

Thus we obtain a center-type solution, i.e., the same qualitative result as in Goodwin's model. The period of the cycle is given approximately by

$$q = 2\pi\ [(\delta g a_2 (1 - b_3)\ \beta_e \lambda_e) / (1 - a_3 b_3)]^{-1/2}. \tag{3.19}$$

Numerical examples show that with reasonable parameter choices the period becomes considerably shorter than in Goodwin's case. Although Goodwin's model has been criticized for producing cycles which are too long to explain business cycles[7)], our modification may serve to demonstrate that a business cycle model with realistic periods might be developed along Goodwinian lines. The parameters underlying

Figure 3, showing a closed solution curve in the λ,β-space, produce cycles of approximately 8.7 years. This curve has been calculated from (3.8) and (3.9) by means of Runge-Kutta-methods.



Parameters: $\mu_1=0$, $\mu_2=1$, $\nu_1=0.04$, $\nu_2=0$, $\gamma_1=0$, $\gamma_2=0$, $\delta=4$, $\sigma=0.55$, $c=0.01$, $c=0.4$, $n=0.01$, $s_\pi=0.24$, $s_w=0.04$, $a_1=0.9$, $a_2=1$, $a_3=0.99$, $b_1=1.9$, $b_2=0$, $b_3=0.6$

Equilibrium values: $\beta_e=0.8931$, $\lambda_e=0.7375$, $\theta_e=0.86580087$, $v_e=2.1$

Initial conditions: $\beta(0)=0.95$, $\lambda(0)=\lambda_e$, $\theta(0)=\theta_e$, $v(0)=v_e$

Number of iterations: 88, step size: 0.1, number of years: 8.8, length of cycle: 8.8

Figure 3 Cyclical Behaviour in the Case of Production Adjustment Only

4. VARIABLE SPEED OF TECHNICAL PROGRESS AND CHANGES IN WORK INTENSITY

In this section we reintroduce the variability of work intensity, i.e., we apply equation (2.13). Moreover, we allow for a variable speed of neutral technical progress by assuming

$$\mu_1 = 0 \ , \ \mu_2 = 1 \ , \ \nu_1 > 0 \ , \ \nu_2 > 0 . \tag{4.1}$$

The latter assumption takes into account the impact of income distribution on mechanization, without affecting technical capital productivity, however. The set of assumptions (4.1) implies that

$$u(\lambda) = \hat{P/\Gamma} = \nu_1 + \nu_2 \lambda \tag{4.2}$$

and

$$\psi(\lambda) = \hat{K/P} = 0 \tag{4.3}$$

and we obtain the following system:

$$\dot{\beta} = [\varphi\ (\beta) + f\ (\lambda) - \nu_1 - \nu_2 \lambda - n]\ \beta \tag{4.4}$$

$$\dot{\lambda} = [\varphi\ (\beta) + \phi_1(\beta) - \nu_1 - \nu_2 \lambda\ ]\ \lambda \tag{4.5}$$

$$\dot{v} = [-f\ (\lambda) + z\ (\lambda, v)]\ v \tag{4.6}$$

$$\theta = 1/(\sigma v)\ . \tag{4.7}$$

As in the case of section 3, the first two equations form an independent subsystem in β and λ, and once again there exists a steady state solution. The equilibrium values turn out to be

$$\beta_e = \left[(1-a_3 b_3)\ \{\delta g(\gamma_1+\nu_1) + \nu_2[\delta(g-s_w)-n]\}\right] / \ [a_2(1-b_3)(\nu_2+\delta g)\ + \gamma_2 \delta g(1-a_3 b_3)] \ +$$
$$+\ \{[a_1(1-b_3) - b_1(1-a_3)]\ (\nu_2+\delta g)\}\ /\ [a_2(1-b_3)(\nu_2+\delta g) + \gamma_2 \delta g(1-a_3 b_3)] \tag{4.8}$$

$$\lambda_e = [\delta(g-s_w) - \nu_1 - n - \gamma_1 + \gamma_2 \beta_e] / (\nu_2 + \delta g) \tag{4.9}$$

$$v_e = [c(1-\lambda_e)]\ /\ \{\delta\ [\ g(1-\lambda_e) - s_w]\} \tag{4.10}$$

$$\theta_e = 1/(\sigma v_e)\ . \tag{4.11}$$

Evaluating the Jacobian at the equilibrium point yields

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where

$$A = \varphi'(\beta_e)\,\beta_e = \gamma_2\beta_e > 0 \tag{4.12}$$

$$B = [\,f'(\lambda_e) - \nu_2]\,\beta_e = -(\delta g + \nu_2)\,\beta_e < 0 \tag{4.13}$$

$$C = [\,\varphi'(\beta_e) + \Phi_1'(\,\beta_e)]\,\lambda_e = [\,\gamma_2 + a_2(1-b_3)\,/\,(1-a_3b_3)]\,\lambda_e > 0 \tag{4.14}$$

$$D = -\nu_2\lambda_e < 0\,. \tag{4.15}$$

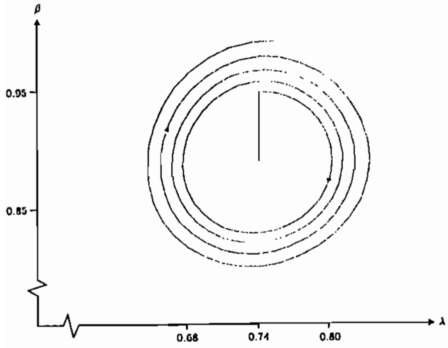
Instead of zero entries in the main diagonal, we have a positive one (A) and a negative one (D). The eigenvalues are given by

$$s_{1,2} = (A + D)/\,2 \pm \{[(A + D)/\,2]^{\,2} + BC - AD\}^{1/2}\,. \tag{4.16}$$

Let us consider the impact of both modifications in turn. If we assume a constant intensity of work, A would be equal to zero. In this case, both eigenvalues are either real and negative, or we have a complex pair with a negative real part. In either subcase the linearized system is locally stable. The opposite happens if we neglect the impact of distribution on mechanization. In this case we have $D = 0$ and the roots will either be real and positive, or complex with a positive real part. In either subcase the linearized system is locally unstable. In all cases $|BC|$ needs to be sufficently large to make the discriminant negative, and thus to produce cyclical solutions.

What will be the combined result of the stabilizing impact of a variable speed of neutral technical progress, and the destabilizing effect of changes in work intensity? We may expect that either one force dominates the other or that they just offset each other. The latter would happen, for instance, if $A + D = 0$ and $BC + A^2 < 0$, in which case an oscillatory movement without trend would occur.

The results derived from the consideration of the linearized subsystem can be sustained by numerical solutions to the non-linear system obtained, for various parameter combinations, by applying a four-step Runge-Kutta algorithm. Figures 4a, b and c represent an unstable, a stable and a borderline situation respectively. While the discriminant of (4.16) is negative for all cases, the real parts of the complex pairs of the roots are positive, negative and (approximately) zero, respectively.



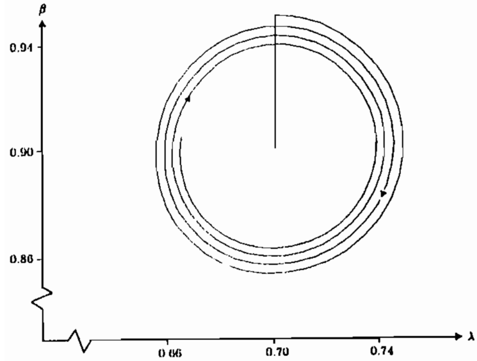
(a)

Parameters: $\mu_1=0$, $\mu_2=1$, $\nu_1=0.02$, $\nu_2=0.03$, $\gamma_1=0.05$, $\gamma_2=0.06$, $\delta=4$, $\sigma=0.55$, $c=0.4$, $n=0.01$, $s_\pi=0.24$, $s_w=0.04$, $a_1=0.9$, $a_2=1$, $a_3=0.99$, $b_1=1.9$, $b_2=0$, $b_3=0.6$

Equilibrium values: $\beta_e=0.89174993$, $\lambda_e=0.73916265$, $\theta_e=0.84814058$, $v_e=2.143727$

Initial conditions: $\beta(0)=0.95$, $\lambda(0)=\lambda_e$, $\theta(0)=\theta_e$, $v(0)=v_e$

Number of iterations: 134, step size: 0.25, number of years: 33.5, length of cycle: 8.4



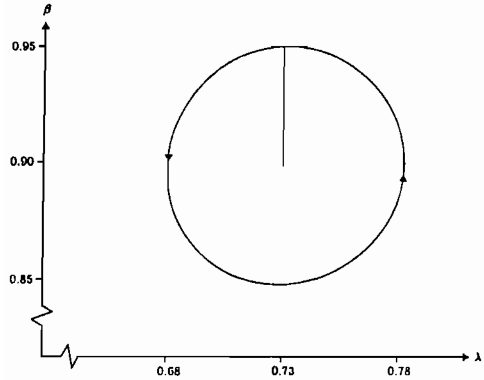
(b)

Parameters: $\mu_1=0$, $\mu_2=1$, $\nu_1=0.02$, $\nu_2=0.04$, $\gamma_1=0.01$, $\gamma_2=0.012$, $\delta=4.2$, $\sigma=0.55$, $c=0.38$, $n=0.01$, $s_\pi=0.24$, $s_w=0.04$, $a_1=0.9$, $a_2=1$, $a_3=0.99$, $b_1=1.9$, $b_2=0$, $b_3=0.6$

Equilibrium values: $\beta_e=0.90048401$, $\lambda_e=0.70201735$, $\theta_e=0.91965929$, $v_e=1.9770167$

Initial conditions: $\beta(0)=0.95$, $\lambda(0)=\lambda_e$, $\theta(0)=\theta_e$, $v(0)=v_e$

Number of iterations: 134, step size: 0.25, number of years: 33.5, length of cycle: 8.9



(c)

Parameters: $\mu_1=0$, $\mu_2=1$, $\nu_1=0.02$, $\nu_2=0.03439163$, $\gamma_1=0.025$, $\gamma_2=0.028$, $\delta=4$, $\sigma=0.55$, $c=0.4$, $n=0.01$, $s_\pi=0.24$, $s_w=0.04$, $a_1=0.9$, $a_2=1$, $a_3=0.99$, $b_1=1.9$, $b_2=0$, $b_3=0.6$

Equilibrium values: $\beta_e=0.89817498$, $\lambda_e=0.73125002$, $\theta_e=0.93023232$, $v_e=1.954546$

Initial conditions: $\beta(0)=0.95$, $\lambda(0)=\lambda_e$, $\theta(0)=\theta_e$, $v(0)=v_e$

Number of iterations: 134, step size: 0.25, number of years: 33.5, length of cycle: 8.5

Figure 4 Cyclical Behaviour
(a) if Work Intensity Effect Dominates
(b) if Distribution-Induced Technical Change Dominates
(c) for a balanced case.

The specific parameter combinations are reported together with the figures. Similar graphs could be drawn for the pairs λ, θ and λ, v.

5. NON-NEUTRAL TECHNICAL PROGRESS

In this section we allow for non-neutral technical progress, while at the same time we apply the simplifying devices (3.1) and (3.5) again; i.e., we neglect changes in work intensity and the influence of capacity utilization on price formation. Parameter restrictions regarding technical progress now become

$$\mu_1 \geqq 0 \ , \ 0 < \mu_2 < 1 \ , \ \nu_1 \geqq 0 \ , \ \nu_2 \geqq 0 . \tag{5.1}$$

If $\mu_1 = 0$, then technical progress will always be capital-using. This corresponds to Marx's assumption of a rising technical composition of capital. If $\mu_1 > 0$, however, an equilibrium solution exhibiting neutral technical progress may exist, while off the equilibrium path technical progress would be either capital saving or labour saving. This corresponds to Kaldor's view. (5.1) implies that

$$u(\lambda) = \hat{P/\Gamma} = \mu_1 + \mu_2\nu_1 + \mu_2\nu_2\lambda \tag{5.2}$$

$$\psi(\lambda) = \hat{K/P} = \nu_1 - \mu_1 - \mu_2\nu_1(1-\mu_2)\lambda \ . \tag{5.3}$$

Under these restrictions we obtain the system

$$\dot{\beta} = [\ f(\lambda) - u(\lambda) - n]\ \beta \tag{5.4}$$

$$\dot{\lambda} = [\ \theta_1(\beta) - u(\lambda)]\ \lambda \tag{5.5}$$

$$\dot{v} = [-f(\lambda) + z(\lambda,v)]\ v \tag{5.6}$$

$$\dot{\theta} = [\ \psi(\lambda) + f(\lambda) - z(\lambda,v)]\ \theta. \tag{5.7}$$

Obviously, this system of four equations is decomposable. Yet, in general, no steady state solution will exist. To see this, consider first equations (5.6) and (5.7). A stationary solution requires that $\psi(\lambda)$ vanishes. From (5.3) we can find the specific value λ^* which would bring that about:

$$\lambda^* = [\ \mu_1 - \nu_1(1-\mu_2)] \ / \ [\ \nu_2(1-\mu_2)]\ . \tag{5.8}$$

Next, consider equations (5.4) and (5.5). This independent subsystem has an equilibrium solution of its own, i.e.,

$$\lambda_e = [\ \delta(g - s_w) - n - \mu_1 - \mu_2\nu_1] \ / \ (\delta g + \mu_2\nu_2) \tag{5.9}$$

$$\beta_e = [a_1(1-b_3) - b_1(1-a_3) + (1 - a_3b_3)(\mu_1 + \mu_2\nu_1 + \mu_2\nu_2\lambda_e)] \ / \ [a_2(1-b_3)]. \tag{5.10}$$

λ_e and λ^* will coincide by change only. Note that with the "Marxian" case, $\mu_1 = 0$. There will be a negative λ^*, except if in addition $\nu_1 = 0$, in which case we would have $\lambda^* = 0$. Without these particular restrictions, however, λ^* can be positive. In this "Kaldorian" case, it can either be greater or smaller than λ_e, while equality between the two would be an exceptionally limiting case.

What kind of behaviour will this system show in the neighbourhood of the equilibrium point? The linearized subsystem (5.4) and (5.5) has the following Jacobian, evaluated at its equilibrium point:

$$\begin{bmatrix} 0 & B \\ C & D \end{bmatrix}$$

where

$$B = [\ f'(\lambda_e) - u'(\lambda_e)]\ \beta_e = -(\delta g + \mu_1 + \mu_2\nu_1 + \mu_2\nu_2\lambda_e)\ \beta_e < 0 \tag{5.11}$$

$$C = \phi_1'(\beta_e)\lambda_e = [\ a_2(1-b_3)/(1-a_3b_3)]\ \lambda_e > 0 \tag{5.12}$$

$$D = -u'(\lambda_e)\ \lambda_e = -\mu_2\nu_2\lambda_e < 0. \tag{5.13}$$

The eigenvalues are given by

$$s_{1,2} = D/2 \pm [BC + (D/2)^2]^{1/2}. \tag{5.14}$$

For positive λ_e the equilibrium is stable because $D < 0$. As $B < 0$ and $C > 0$, we get $BC < 0$. Therefore, damped oscillations may occur.

Now consider equation (5.6). Obviously there is a stationary value for the capital coefficient, such that

$$v_e = c(1-\lambda_e) \,/\, f(\lambda_e). \tag{5.15}$$

If the wage share has closely approximated its equilibrium value the movement of the capital coefficient can be approximated by

$$\dot{v} \simeq -f(\lambda_e)(v-v_e). \tag{5.16}$$

It follows that the capital coefficient approaches its equilibrium value. But then we can conclude from (5.7) that

$$\lim_{t\to\infty} \hat{\theta} = \psi(\lambda_e). \tag{5.17}$$

Recalling that $\psi'(\lambda) > 0$, we find

$$\lambda_e \gtreqless \lambda^* \leftrightarrow \psi(\lambda_e) \gtreqless 0 \leftrightarrow \hat{\theta} \gtreqless 0. \tag{5.18}$$

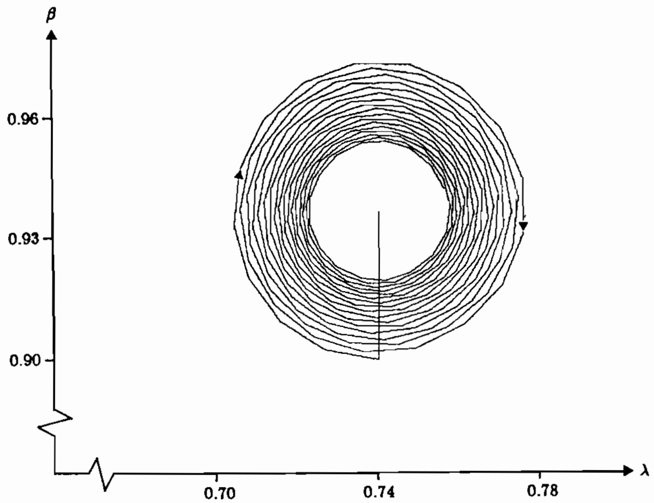
Thus, there is only one special case, $\lambda^* = \lambda_e$, in which capacity utilization will approach a steady state value in the long run. For $\lambda_e > \lambda^*$, and for the "Marxian" case in particular, we obtain a long-term increase in capacity utilization. A downward trend will be established in the opposite situation, i.e., for $\lambda_e < \lambda^*$. Moreover, from (3.16) we have

$$\hat{\theta} = -\hat{v} - \hat{\sigma}. \tag{5.19}$$

As the capital coefficient approaches a steady state value, we obtain for the growth rate of technical capital productivity

$$\lim_{t\to\infty} \hat{\sigma} = -\lim_{t\to\infty} \hat{\theta} = -\psi(\lambda_e). \tag{5.20}$$

The economic interpretation is as follows: Whenever the equilibrium wage share happens to be "too high" (that is, whenever λ_e exceeds λ^*) it causes a long run fall of technical capital productivity growth by stimulating "too much" mechanization. However, this fall will not induce a rise of the capital coefficient. Instead, the latter approaches a steady state value, which is made possible by a compensatory increase of capacity utilization. This somewhat strange result is the consequence of a missing feedback from capacity utilization to the wage share, i.e., of our simplifying assumption (3.6). If, however, high capacity utilization rates gave rise to price increases which could not be fully matched by wage rises, then the equilibrium wage share might be pushed down to a level coinciding with λ^*. We shall see that this may happen in the general case. One could argue that this case does not make much economic sense. Considering numerical examples, however, one finds that the trend component of capacity utilization is rather weak compared with its cyclical dynamics. In the example visualized in Figures 5 and 6, it takes about 200 years to induce the rise



Parameters: μ_1=0.018, μ_2=0.5, ν_1=0.015, ν_2=0.03, γ_1=0, γ_2=0, δ=4,
n=0.01, c=0.4, s_π=0.24, s_w=0.04, a_1=0.9, a_2=1, a_3=1, b_1=1.9, b_2=0, b_3=0.6

Equilibrium values: β_e=0.93662577, λ_e=0.74171779, θ_e=+∞, v_e=2.2157895

Initial conditions: $\beta(0)$=0.9, $\lambda(0)=\lambda_e$, $\theta(0)$=0.8, $v(0)=v_e$

Number of iterations: 280, step size: 0.5, number of years: 140,
length of cycle: 8.4

Figure 5 Non-neutral Technical Progress. The Movements of the Labour Share and the Degree of Employment of capacity utilization from an initial value of 0.8 to 0.9. This does not seem to render the model nonsensical for a medium-term analysis.

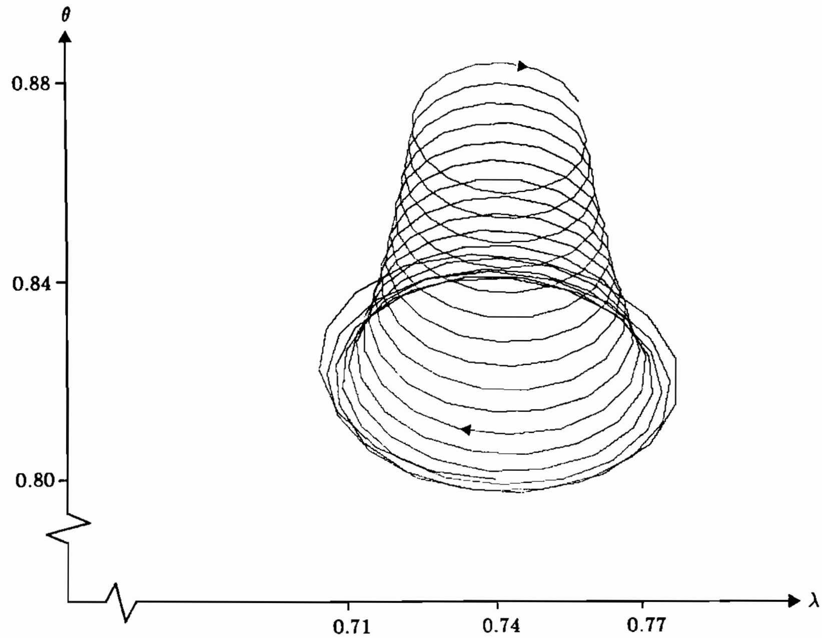


Figure 6 Non-neutral Technical Progess. The Movements of the Labour Share and Capacity Utilization (Parameters as in Figure 5)

6. THE GENERAL MODEL RECONSIDERED

After having analyzed a sequence of simplified cases, we shall now return to the general model of section 2. While it is not decomposable, it has a stationary solution. From (2.51) and (2.52) it follows that for a steady state solution to the system to exist, $\psi(\lambda)$ has to vanish in equilibrium. Therefore we have

$$\lambda_e = [\mu_1 - \nu_1(1-\mu_2)] \,/\, [\nu_2(1-\mu_2)] \tag{6.1}$$

which, of course, is identical with λ^* from (5.8). Next, we find from (2.49) that

$$\beta_e = \{\gamma_1 + n + \mu_1/(1-\mu_2) - \delta[g(1-\lambda_e) - s_w]\} \,/\, \gamma_2. \tag{6.2}$$

In equilibrium, the capital coefficient will be

$$v_e = [c(1-\lambda_e)] \,/\, \{\delta[g(1-\lambda_e) - s_w]\} \,. \tag{6.3}$$

Finally, we calculate from (2.50) that

$$\theta_e = \big[[\varphi(\beta_e) + \phi_1(\beta_e) - u(\lambda_e)] \,/\, [b_2(1-a_3)]\big]\,(1 - a_3 b_3). \tag{6.4}$$

This time we have to consider the full-scale model to obtain any definite results about its motion off the equilibrium path. Linearized around the equilibrium point, the system has the following Jacobian:

$$\begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ 0 & a_{32} & 0 & a_{34} \\ 0 & a_{42} & 0 & a_{44} \end{bmatrix}$$

where

$$\begin{aligned}
a_{11} &= \varphi'(\beta_e)\beta_e && = \gamma_2\beta_e > 0 \\
a_{12} &= [f'(\lambda_e) - u'(\lambda_e)]\,\beta_e && = -(\delta g + \mu_2\nu_2)\,\beta_e < 0 \\
a_{21} &= [\varphi'(\beta_e) + \phi_1'(\beta_e)]\,\lambda_e && = [\gamma_2 + a_2(1-b_3)/(1-a_3b_3)]\,\lambda_e > 0 \\
a_{22} &= -u'(\lambda_e)\,\lambda_e && = -\mu_2\nu_2\lambda_e < 0 \\
a_{23} &= -\phi_2'(\theta_e)\,\lambda_e && = -b_2(1-a_3)\,\lambda_e/(1-a_3b_3) < 0 \\
a_{32} &= [\psi'(\lambda_e) + f'(\lambda_e) - \partial z/\partial\lambda_e]\,\theta_e && = [\,\nu_2(1-\mu_2) - \delta g + c/v_e]\,\theta_e \\
a_{34} &= -(\partial z/\partial v_e)\,\theta_e && = [\,c(1-\lambda_e)/v_e^2]\,\theta_e > 0 \\
a_{42} &= [-f'(\lambda_e) + \partial z/\partial\lambda_e]\,v_e && = (\delta g - c/v_e)v_e \\
a_{44} &= (\partial z/\partial v_e)v_e && = -c(1-\lambda_e)/v_e < 0.
\end{aligned}$$

The characteristic equation reads

$$s^4 + c_1 s^3 + c_2 s^2 + c_3 s + c_4 = 0 \tag{6.5}$$

where

$$c_1 = -(a_{11} + a_{22} + a_{44})$$

$$c_2 = a_{11}a_{22} + (a_{11} + a_{22})a_{44} - a_{12}a_{21} - a_{23}a_{32}$$

$$c_3 = (a_{12}a_{21} - a_{11}a_{22} + a_{23}a_{32})a_{44} + (a_{11}a_{32} - a_{34}a_{42})a_{23}$$

$$c_4 = a_{11}a_{23}(a_{34}a_{42} - a_{32}a_{44}).$$

In general, we shall obtain four non-zero eigenvalues which can, in principle, be calculated for numerial examples. If we only want to check whether a specific example is locally stable or not, it is more convenient to apply the Liénard-Chipard conditions. As numerical methods are available to solve the non-linear system, we prefer this approach and again apply the Runge-Kutta-methods.

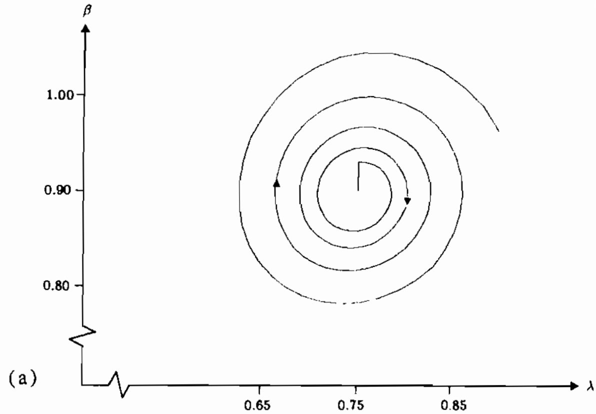
Consider, for example, the parameter constellation

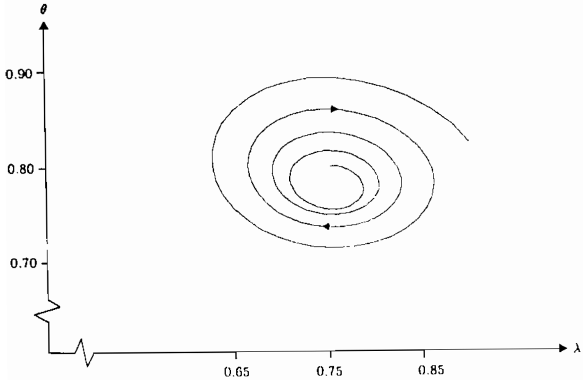
$\mu_1 = 0.012$	$\mu_2 = 0.60$	
$\nu_1 = 0.015$	$\nu_2 = 0.02$	
$\gamma_1 = 0.09$	$\gamma_2 = 0.10$	
$s_\pi = 0.24$	$s_w = 0.04$	
$c = 0.40$	$\delta = 4$	$n = 0.01$
$a_1 = 0.865$	$a_2 = 1$	$a_3 = 0.90$
$b_1 = 0.190$	$b_2 = 0.25$	$b_3 = 0.50$

It gives rise to the following equilibrium values of our most important variables:

$$\beta_e = 0.90 \qquad \lambda_e = 0.75 \qquad \theta_e = 0.80 \qquad v_e = 2.50.$$

Along the steady state path, prices and money wages will rise by 5% and 8% p.a., respectively, while net output and the capital stock increase by 4% p.a.. The intensity of work remains constant along the steady state path, while labour productivity rises by 3%. Figures 7a, b and c show the development of the most relevant variables over





(b)

Parameters: $\mu_1=0.012$, $\mu_2=0.6$, $\nu_1=0.015$, $\nu_2=0.02$, $\gamma_1=0.09$, $\gamma_2=0.1$, $\delta=4$, $c=0.4$, $n=0.01$, $s_\pi=0.24$, $s_w=0.04$, $a_1=0.865$, $a_2=1$, $a_3=0.9$, $b_1=0.19$, $b_2=0.25$, $b_3=0.5$

Equilibrium values: $\beta_e=0.9$, $\lambda_e=0.75$, $\theta_e=0.8$, $v_e=2.5$

Initial values: $\beta(0)=0.93$, $\lambda(0)=\lambda_e$, $\theta(0)=\theta_e$, $v(0)=v_e$

Number of iterations: 144, step size: 0.25, number of years: 36, length of cycle: 8.6

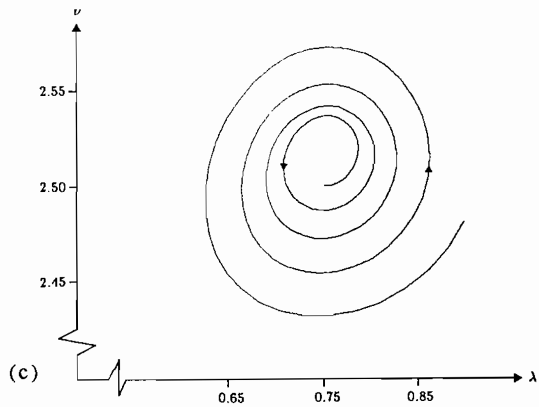


Figure 7 Movements of the Labour Share and
(a) the Degree of Employment in the General Case
(b) Capacity Utilization in the General Case
(c) the Capital Coefficient in the General Case

time, under the assumption of an initial displacement of the rate of employment of +0.02 from its equilibrium value. Obviously, an explosive cyclical motion is obtained, the length of the cycle being approximately 8.6 years.

Yet the explosive character of the cycles is not inevitable. For the pair $\gamma_1 = 0.0009$ and $\gamma_2 = 0.001$, the equilibrium values remain the same but the system turns out to be mildly stable.

It should be noted that similar examples could be constructed assuming a small steady state fall in work intensity (or working hours), which would reduce (net) labour productivity growth (per head) accordingly.

Our calculations lead to the conclusion that Goodwin's growth cycle model is capable of considerable extensions which may enrich its empirical flavour. On the other hand, it must be recognized that these extensions spoil the characteristic feature of a closed solution curve which made Goodwin's original model so attractive. It would be more appealing to present a set of modifications which produced limit cycle solutions. However, those known to us seem to rely too heavily on special assumptions.

7. CONCLUSIONS

We have discussed some modifications of Goodwin's growth cycle model. If only an adjustment process of production with respect to demand is taken into account, the oscillations of the wage share and the rate of employment will exhibit closed orbits in the phase space. Therefore, an important characteristic of Goodwin's original model is preserved. Moreover, it can be shown that the length of the cycle may become much shorter than in the original model, thereby adding to its realism. Introducing the variability of work intensity renders the model cyclically unstable, while permitting a distribution-induced change in the speed of technical progress serves as a stabilizing force. The net outcome of both impacts can work in either direction, including the special case of balance between the two opposing forces.

The introduction of capital-augmenting technical progress, together with production adjustment, precludes a steady state solution for the system as a whole. While the rate of employment, the labour share and the capital coefficient approach positive equilibrium values (under reasonable parameter restrictions), capital utilization will show a damped cyclical movement around a trend.

The full-scale model - including an additional feedback from capacity utilization to price formation and real wage changes – is shown to be capable of a steady state solution and cyclical movements of all variables in disequilibrium. Numerical examples suggest that unstable cyclical motions may be expected, although constellations exist which render the system stable.

Our modified model could be developed further. For the time being, not all economically meaningful combinations amongst the present set of modifications have been analysed separately. A more detailed investigation may enhance our understanding. Moreover, the present modifications could be implemented differently, for example, by using alternative types of functions. Also, one could take into account the impact on productivity growth of changes in capacity utilization which arise from the phenomenon of labour hoarding. This procyclical impact on productivity could weaken or even reverse the countercyclical effect of changes in work intensity. Another modification may be derived from the principle of "absolute overaccumulation" (cf. Glombowski, Krüger 1985). This principle addresses the problem of regime switches in capital accumulation: if profits fail to rise at a growth rate acceptable to investors, the speed of capital accumulation drops sharply, thereby serving as a kind of boundary to unstable motions. It might be fruitful to apply this regime switch in a more general model of capital accumulation. Finally, we think it worthwhile to substitute our nominal (augmented) wage equation by a bargaining-theoretical approach. This might improve the behavioural foundation of our approach to the theory of cyclical growth.

ACKNOWLEDGEMENT

We wish to thank an anonymous referee for helpful comments and constructive criticism.

NOTES

[1] See Goodwin (1972), Goodwin, Krüger, and Vercelli (1984), Glombowski, Krüger (1983a) and (1985).

[2] Goodwin states: "It has long seemed to me that Volterra's problem of the symbiosis of two populations - partly complementary, partly hostile - is helpful in the understanding of the dynamical contradictions of capitalism, especially when stated in a more or less Marxian form." (Goodwin 1972, p. 445).

[3] This effect has been considered in another paper. See Glombowski, Krüger, (1983b).

[4] As the net product need not necessarily be sold completely, profits may partly exist in the form of inventories. Major theoretical problems are not necessarily involved here if, on the average, inventory growth does not outstrip product growth. Whether the latter condition is met depends, of course, on the solution of the model.

[5] Underconsumptionists are invited to analyse the consequences of the opposite assumption.

[6] In some European countries, e.g., Italy and the Netherlands, wage indexation has been practised for considerable periods of time. This may render $a_3 = 1$ a realistic assumption.

[7] Cf. Atkinson (1969).

APPENDIX: LIST OF VARIABLES

β	rate of employment
λ	wage share
θ	capacity utilization
v	capital coefficient
P	maximum output
Y	net product, income
Π	profits
K	capital stock
S	savings
A	labour supply
Γ	efficient labour requirements
L	labour input
y	labour productivity
h	index of labour inefficiency

w	nominal wage rate
p	price level
ω	real wage rate
σ	technical capital productivity
γ	technical labour productivity

REFERENCES

Atkinson, A.B., 'The Time Scale of Economic Models: How Long is the Long Run? Review of Economic Studies , 36, 1969, pp. 137-152.

Bauer, O., The Accumulation of Capital, History of Political Economy, 18, 1986, pp. 88-110 (German original 1913).

Bergstrom, A.R., The Construction and Use of Economic Models, The English Universities Press Ltd., London, 1967.

Bowles, S., Gordon, D.M., Weisskopf, Th. E., Beyond the Waste Land: A Democratic Alternative to Economic Decline, Anchor Press/Doubleday, Garden City, New York, 1983.

Glombowski, J., Krüger, M., On the Rôle of Distribution in Different Theories of Cyclical Growth, Research Memorandum 132, Department of Economics, Tilburg University, 1983a.

Glombowski, J., Krüger, M., On Changes of Labour Productivity Growth in a Model of Cyclical Growth, paper presented to the Elfde Nationale Onderzoekdag Economie, Tilburg, 1983b.

Glombowski, J., Krüger, M., Absolute Overaccumulation of Capital in a Model of Cyclical Growth, unpublished paper, Osnabrück, 1985.

Goodwin, R.M., A Growth Cycle, in: Hunt, E.K., Schwartz. J.G., (eds), A Critique of Economic Theory; Selected Readings, Penguin, Harmondsworth, 1972, pp. 442-449.

Goodwin, R.M., Krüger, M., Vercelli, A., (eds), Nonlinear Models of Fluctuating Growth, Springer, Berlin et al., 1984.

Itoh, M., Value and Crisis, Verso, London, 1980.

Kaldor, N., A Model of Economic Growth, (1957), in: Kaldor, N., 1960, Essays on Economic Stability and Growth, Duckworth, London, 1960.

Kaldor, N., Essays on Economic Stability and Growth, Duckworth, London, 1960.

Keynes, J.M., The Collected Writings of John Maynard Keynes, Vol. VII, The General Theory of Employment, Interest, and Money, Macmillan Press, London and Basingstoke, 1973.

Marx, K., Capital. A Critical Analysis of Capitalist Production, Vol. I, Lawrence and Wishart, London, 1974.

Phillips, A.W., Stabilization Policy in a Closed Economy, Economic Journal 64, 1954, pp. 290-323.

Solow, R.M., Stiglitz, J.E., Output, Employment, and Wages in the Short Run, Quarterly Journal of Economics 82, 1968, pp. 537-560.

Sweezy, P.M., The Theory of Capitalist Development, New York, 1942.

Volterra, V., Lecons sur la Théorie mathématique de la Lutte pour la Vie, Paris, 1931.

Chapter 12

A Multisector Model of the Trade Cycle

A. MEDIO

1. The main purpose of this paper is to generalize to the multidimensional case certain results established for aggregate multiplier-accelerator models. The starting point of our investigation will be the well known dynamic input-output model:

$$x = Ax + B\dot{x} \qquad (1)$$

where $x \in R^n$ is a column vector indicating the levels of activity of the various sectors; $A \in R^{nxn}$ and $B \in R^{nxn}$ are matrices indicating, respectively, the flow and the stock input-output relations.

For n = 1, system (1) collapses into the aggregate multiplier-accelerator model, where x is income, the scalars A and B are, respectively, the propensity to consume and the acceleration coefficient. System (1) implicitly assumes that demand (desired supply) and actual supply, as well as desired and actual investment, are always in equilibrium. Moreover, since A and B are constant, the model is linear.

Subsequently, we shall drop the former assumption, allowing for discrepancies between desired and actual quantities, and replace the equilibrium conditions with adjustment-mechanisms of the <u>tâtonnement</u> type.

The linearity assumption will be removed partially, by considering variable accelerator coefficients. Thus, equation (1) will be re-interpreted as a linearized version of an underlying nonlinear model.

We shall then perform a stability analysis of the system, studying the dependence of its solutions on the parameters (input-output coefficients, lags). In particular, the conditions for periodic solutions of the system, which correspond to persistent oscillations of the levels of activity of the economy will be established. In so doing, we shall make use of some techniques of analysis recently introduced into mathematical economics, namely vector Lyapunov functions and bifurcation theory.

2. The first step in this direction will be a brief discussion of a simpler two-dimensional model. Certain established results will be reviewed, and the existence and stability of limit cycles will be proved by means of the Poincaré-Andronov-Hopf bifurcation theorem in R^2 [1].

If only one sector exists in the economy, and if we re-name the variables to make the reference to the business cycles literature more familiar, the multiplier-accelerator model can be written thus:

$$Y = C + I$$
$$C = cY \qquad c = 1 - s$$
$$I = v\dot{Y} \qquad 0 < c < 1$$

or

$$Y = cY + v\,\dot{Y} \tag{2}$$

where Y, C and I indicate, respectively, global income, consumption and investment; c is the propensity to consume, s the propensity to save and v the accelerator coefficient.

Let us now introduce two simple exponential lags, one between demand and supply, and the other between desired and actual investment.

If T_y and T_i, are the positive lengths of the lags and D = (d/dt), t being time, we can rewrite (2) as

$$(T_yD + 1)Y = C + I$$
$$C = cY$$
$$(T_iD + 1)I = v\dot{Y} \qquad (3)$$

whence

$$T_yT_i\ddot{Y} + (T_y+sT_i-v)\dot{Y} + sY = 0 \qquad (4)$$

or, in a state space form, putting $X = \dot{Y}$,

$$\begin{bmatrix} \dot{Y} \\ \dot{X} \end{bmatrix} = \begin{bmatrix} 0, & 1 \\ -[T_yT_i]^{-1}s, & [T_yT_i]^{-1}[v-(T_y+sT_i)] \end{bmatrix} \begin{bmatrix} Y \\ X \end{bmatrix}. \qquad (5)$$

System (5) has only one equilibrium point at which $\dot{X} = \dot{Y} = 0$, i.e. the origin of the co-ordinates $X = Y = 0$). The characteristic roots of (5), μ_1, μ_2, are equal to

$$\mu_{1,2} = 1/2\{[T_yT_i]^{-1}[v-(T_y+sT_i)] \pm \{[[T_yT_i]^{-1}[v-(T_y+sT_i)]]^2 - 4[T_yT_i\}^{-1}s\}^{1/2}\}.$$

As $-[T_yT_i]^{-1} s < 0$, $\mu_1\mu_2 > 0$ and the equilibrium cannot be a saddle point. Re $\mu_{1,2} \lesseqgtr 0$, i.e. the equilibrium will be stable or unstable, according to whether

$$v - (T_y + sT_i) \lesseqgtr 0.$$

This means that strong accelerators work for instability, whereas the greater the propensity to save, and the longer the lags, the more likely it will be for the system to be stable. The roots will be complex or real - i.e. the motion will be oscillatory or not - according to whether

$$[v - (T_y + sT_i)]^2 \lesseqgtr 4\,s.$$

3. Let us now consider a nonlinear version of (5). In particular, let us assume that

v, the acceleration coefficient, is no longer constant but a function of the level of income.

In economic terms, v can be interpreted as the (incremental) desired capital-output ratio, so that we can write:

$$v(Y) = [dK(Y)/dY] \ , \qquad (6)$$

K being the capital stock.

Following the vast literature on the subject, we postulate that the function K(Y) is of a "saturation" type, i.e.: that the desired capital stock increases with income within a certain range above or below the normal level of income, but flattens out at exceptionally high or exceptionally low levels. If we expand the function K(Y) in a Taylor series around Y=0,

$$K(Y) = K(0) + (dK/dY)Y + [(d^2K/dY^2)/2!]Y^2 + ...,$$

where the quantity

$$\bar{v} = v(o) = (dK/dY)\,|_{Y=0}$$

measures the slope of the curve K(Y) at the origin of the coordinates, and can be interpreted as the equilibrium or "normal" accelerator.

The "complete" nonlinear model can now be written thus:

$$\begin{bmatrix} \dot{Y} \\ \dot{X} \end{bmatrix} = \left\{ \begin{bmatrix} 0, & 1 \\ -[T_yT_i]^{-1}s, & [T_yT_i]^{-1}\ [v-(T_y+sT_i)] \end{bmatrix} + \begin{bmatrix} 0, & 0 \\ 0, & f(Y) \end{bmatrix} \right\} \begin{bmatrix} Y \\ X \end{bmatrix} \qquad (7)$$

where $f(Y) = 0(Y^2)$ is the nonlinear component of (7). The "complete" system (7) also has a unique equilibrium point at $\dot{Y}=Y=0$ and, in general, local stability may be ascertained by considering the linear part only. We shall now consider the system as depending on the parameter $\bar{v}$, and shall study the system as its behaviour varies qualitatively when $\bar{v}$ is changed.

Suppose we start with a weak "normal" accelerator - i.e. $\bar{v} < (T_y+sT_i)$. In this case, stability of equilibrium of the system (7) can be predicted by the fact that the characteristic roots of its linear component have negative real parts. At the critical point

$$\bar{v}_0 = T_y + sT_i,$$

however, the two characteristic roots, μ_1 and μ_2, will be purely imaginary with zero real parts.

In this case, which corresponds to the equilibrium point being of a centre type, the linear part of the system does not give us a qualitatively accurate picture of the "complete" nonlinear system. In particular, it cannot be used to assess the stability of the equilibrium point, and it fails to detect non-trivial periodic solutions i.e. limit cycles. Existence of the latter can be proved, however, using the Hopf bifurcation theorem.

We omit a full statement of the theorem here [2] , and only retain the following essential conditions for a system like (7), to have periodic solutions, depending on a parameter $\bar{v}$, namely:

(i) there exists an isolated stationary equilibrium point;
(ii) there exists a pair of complex conjugate characteristic roots, μ and $\bar{\mu}$ of the linear part of the system such that, for $\bar{v}=\bar{v}_0$,

$$\begin{aligned} &\mathrm{Re}\,\mu = 0 \\ &\mathrm{Im}\,\mu \neq 0 \qquad (8) \\ &(\partial \mathrm{Re}\,\mu(\bar{v})/\partial\bar{v}) \neq 0; \end{aligned}$$

(iii) the remaining characteristic roots have strictly negative real parts.

In our case, at $\bar{v} = \bar{v}_0$, besides Re $\mu = 0$, we have

$$\text{Im } \mu = \{[T_y T_i]^{-1} s\}^{1/2} = \Omega > 0$$

$$[\partial \text{Re } \mu \ (\bar{v}) / \partial \bar{v}] = (1/2)[T_y T_i]^{-1} > 0$$

and the condition (iii) is irrelevant here. Therefore the conditions (8) exist, and we can conclude that when the accelerator is increased _vis-a-vis_ saving ratio and lags, and the system loses its stability, self-oscillations will occur. Their amplitude will be of order $|\bar{v} - \bar{v}_0|^{1/2}$ and their period will be of order $(2\pi / \Omega)$. These solutions will be stable if the limit cycle exists for $\bar{v} > \bar{v}_0$ but unstable in the opposite case.

4. To assess the uniqueness and stability of limit cycles ascertained by means of the Hopf bifurcation is usually quite a formidable task, even in small dimension systems, and we shall not attempt it here. For two-dimensional systems, on the other hand, existence, uniqueness, and stability of limit cycles can be studied by means of simpler and more efficient theorems [3]. Instead, our next step will be to extend our analysis to the multidimensional case, and ascertain to what extent the results just obtained may be generalized. For this purpose, the Hopf bifurcation is one of the few reliable methods, so we shall make use of it.

Introducing production and investment lags of a simple exponential type into system (1) in R^n we have:

$$(T_j^{(y)} D + 1) x_j = x_j^D$$

$$x_j^D = x_j^C + x_j^I$$

$$x_j^C = \sum_{i=1}^{n} a_{ij} x_i$$

$$(T_j^{(i)}D+1)x_j = \sum_{i=1}^{n} b_{ji}\,\dot{x}_i \qquad j=1,2,\ldots,n;$$

where $T_j^{(y)}$ and $T_j^{(i)}$ indicate, respectively, the lengths of the production and investment lags; x_j^D, x_j^C, x_j^I, indicate demand, consumption and investment respectively, all with regard to commodity j; the meanings of the other symbols are the same as before.

For the entire system, in matrix form, we thus have the following equation

$$[T^{(y)}T^{(i)}]\,\ddot{x} + \{T^{(i)} + [T^{(i)}]\,[I-A] - B\}\dot{x} + [I-A]\,x = 0, \qquad (9)$$

where $T^{(y)}$ and $T^{(i)}$ are (nxn) positive diagonal matrices whose elements are, $T_j^{(y)}$ and $T_j^{(i)}$, respectively, and of course I is the (nxn) identity matrix. By introducing two vectors z_1, $z_2 \in R^n$ such that

$$z_1 = \dot{x}$$

$$z_2 = x,$$

our system can be written as:

$$\dot{z} = Gz \qquad (10)$$

where $z = (z_1, z_2) \in R^{2n}$; $G \in R^{2nx2n}$ can be partitioned thus:

$$\begin{bmatrix} G_2 & G_1 \\ I & 0 \end{bmatrix}$$

where 0, I, $G_1, G_2 \in R^{nxn}$ and

$$G_1 = [T^{(y)}T^{(i)}]^{-1}[I-A]$$

$$G_2 = -[T^{(i)}]^{-1} - [T^{(y)}]^{-1}[I-A] + [T^{(y)}T^{(i)}]^{-1} B.$$

Evidently, system (9)-(10) has only one equilibrium point, namely $\dot{z} = z = 0$, the origin of the co-ordinates. It is also readily found that, under reasonable assumptions, det $G > 0$, independently of the accelerator matrix B. Indeed, by making use of the partitioning of G and of a known theorem of matrix algebra [4] , we have

$$\det G = \det(-G_1) = \det([T^{(y)}T^{(i)}]^{-1}[I-A] = \det([T^{(y)}T^{(i)}]^{-1}) \cdot \det(I-A).$$

$T^{(y)}$ and $T^{(i)}$ are diagonal matrices with positive elements, and so are their inverses. Therefore $\det([T^{(y)}T^{(i)}]^{-1}) > 0$. On the other hand, if the flow input-output matrix A possesses the usual viability properties, det (I-A) > 0, and this completes the proof. Considering that $G \in R^{2nx2n}$, it follows that the characteristic roots with negative real parts, as well as those with positive real parts, must be in even numbers. A most important corollary of this proposition is that, in general, <u>when the system moves from stability to instability, there must be two complex conjugate roots whose real parts simultaneously cross the imaginary axis</u>.

5. We shall now try to generalize the perturbation analysis performed in the introductory sections. We shall again choose the acceleration coefficient as the controlling parameter. However, in the multisector model we have n^2 such coefficients, and little hope of reaching meaningful conclusions unless some drastic simplification is introduced. We shall therefore put

$$B = \alpha B_o$$

where B_o is a non-negative [5] constant matrix, and $\alpha > 0$ the scalar controlling parameter. In other words, by varying α we change the scale of the accelerators without altering their structure.

It is now possible to show that when α is small - i.e. when the accelerators are weak <u>vis-à-vis</u> the saving coefficients (the numbers $[1-a_{ij}]$) and the lengths of the lags - the system (10) will be stable, as was the case in the aggregate model.

For this purpose, let $\alpha=0$ (and therefore B=0) and let us define the matrices:

$$-[T^{(y)}]^{-1}[I-A] \equiv H$$
$$[T^{(i)}]^{-1} \equiv T.$$

In this case we have

$$G_1 = TH$$

$$G_2 = H-T.$$

Consider now that the auxiliary equation of G

$$\det(\mu I - G) = 0$$

(μ being a characteristic root) can be written as

$$\det(\mu^2 I - \mu G_2 - G_1) = 0$$

or

$$\det(\mu^2 I - \mu(H-T) - TH) = 0. \qquad (11)$$

Equation (11) can be easily factorized thus:

$$\det(\mu-H) \cdot \det(\mu+T)=0.$$

Hence the 2n characteristic roots of G are as follows:

(i) n characteristic roots are equal to $(-t_i^{-1})$, the latter being an element of the matrix $-[T^{(i)}]^{-1}$, and therefore negative;
(ii) the remaining n characteristic roots are those of the matrix $H \equiv -[T^{(y)}]^{-1}[I-A]$. $T^{(y)}$ is a positive diagonal matrix and, under the customary assumptions about the matrix A, $-[I-A]$ is D-stable [6] . Therefore, H is stable and all its characteristic roots have negative real parts. Q.E.D.

6. We now turn to considering the case in which $\alpha > 0$ and therefore $B \geq 0$; i.e. some

or all of the accelerators are positive.

The first question we want to answer is: what are the conditions under which, for $\alpha > 0$, G retains its stability? In so doing, we shall apply certain recent techniques of analysis of large scale dynamic systems, in particular, the vector Lyanupov functions method [7].

Let us call S the system described by equation (10). We can again partition the state vector $z \in R^{2n}$ into vector components $z_1, z_2 \in R^n$, which yields the system in the form

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} G_2 & G_1 \\ I & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \tag{12}$$

where now $G_2 = H + \alpha\ [T^{(y)}T^{(i)}]^{-1}\ B_0\ - T$. This can be conveniently transformed into

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \left\{ \begin{bmatrix} (H+\alpha B^* & T \\ -\alpha B^* & -T \end{bmatrix} \right\} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \tag{13}$$

where $B^* \equiv [T^{(y)}T^{(i)}]^{-1}B_0$. Systems (12) and (13) have the same auxiliary equation

$$\det(\mu^2 I - \mu[H + \alpha B^* - T] - TH) = 0$$

and therefore they have the same characteristic roots.

If we identify the state vectors z_1, z_2 with two subsystems S_1, and $S_2 \in R^{n x n}$ we can write:

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \left\{ \begin{bmatrix} H & 0 \\ 0 & -T \end{bmatrix} + \begin{bmatrix} \alpha B^* & T \\ -\alpha B^* & 0 \end{bmatrix} \right\} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \tag{14}$$

where the first set of submatrices on the R.H.S. of (14) can be taken to describe the "free" dynamics of the two subsystems, when the feedbacks of each subsystem on itself (i.e. the matrix αB^*) as well as the interconnections between the two subsystems (the matrices T and $-\alpha B^*$) are ignored.

We already know that H and -T are stable matrices and so are the "free" dynamics of S_1 and S_2.

To find stability conditions for the interconnected system $S \in R^{2nx2n}$ by means of the Lyapunov vector functions method, we want to construct a system of differential inequalitites

$$\dot{v} \leqq Wv \tag{15}$$

where $v \in R^2$ is a vector whose elements are Lyapunov functions and $W \in R^{2x2}$ is a matrix to be defined.

To construct suitable functions v, consider again the "free" subsystem S_1 and S_2 and the related equations

$$\dot{z}_1 = Hz_1 \tag{16}$$

$$\dot{z}_2 = -Tz_2. \tag{17}$$

As H and −T are stable matrices, we can write the following equations:

$$H^TP_1 + P_1H = -Q_1 \tag{18}$$

$$(-T)P_2 + P_2(-T) = -Q_2 \tag{19}$$

where P_i, $Q_i \in R^{nxn}$ are symmetric positive definite (s.p.d.) matrices, and the superscript T indicates transpose. A known theorem [8] states that if H and (-T) are stable, for any given s.p.d. matrix Q_i (i=1,2) there exists one (and only one) [9] s.p.d. matrix P_i which satisfies equations (18)-(19). We shall fix

$Q_1 = I$

$Q_2 = 2T$

(from which, of course, $P_2 = I$).

Let us now define the following Lyapunov functions v_i: $fxR^n \rightarrow R_+$ (i=1,2):

$$v_1(z_1) = < P_1 z_1, z_1 > \quad (20)$$

$$v_2(z_2) = < P_2 z_2, z_2 > \quad (21)$$

where $<.>$ indicates the scalar product operation. From (20) and (21) we have

$$\text{grad } v_1 = 2P_1 z_1 \quad (22)$$

$$\text{grad } v_2 = 2P_2 z_2. \quad (23)$$

Let us now take the total time derivative of the functions v_i along the solutions of the <u>interconnected subsystems</u> (13) - (14),

$$\dot{v}_1 = (\text{grad } v_1)^T \dot{z}_1 = (\text{grad } v_1)^T \{[H+\alpha B^*] z_1 + Tz_2) \quad (24)$$

$$\dot{v}_2 = (\text{grad } v_2)^T \dot{z}_2 = -(\text{grad } v_2)^T [\alpha B z_1 + T z_2] \quad (25)$$

whence, by making use of (22) and (23),

$$\dot{v}_1 = 2[<P_1 z_1, Hz_1> + \alpha <P_1 z_1, B^* z_1> + <P_1 z_1, Tz_2>] \quad (24.1)$$

$$\dot{v}_2 = -2[\alpha <z_2, B^* z_1> + <z_2, Tz_2>]. \quad (25.1)$$

Considering now that P_1 and P_2 are symmetric matrices, and making use of the Rayleigh quotient, we can write the following estimates:

$$\mu_m(P_i) \,||z_i||^2 \leq v_i(z_i) \leq \mu_M(P_i) \,||z_i||^2 \quad i=1,2 \quad (26)$$

where, now and in what follows, $\mu_M(M)$ and $\mu_m(M)$ indicate, respectively, the largest and the smallest characteristic root of the matrix M, and $\| x\| \equiv \langle x, x\rangle^{1/2}$ is the Euclidian norm of the vector x. As $P_2 = I$, we immediately get

$$v_2 = \| z_2\|^2.$$

Moreover,

$$\langle P_1 z_1, Hz_1\rangle = -(1/2)\langle z_1, Q_1 z_1\rangle = -(1/2)\| z_1\|^2 \tag{27}$$

$$t_m\| z_2\| = \mu_m(T)\| z_2\|^2 \leqq \langle z_2, Tz_2\rangle \leqq \mu_M(T)\| z_2\|^2 = t_M\| z_2\|^2, \tag{28}$$

where t_m and t_M are, respectively, the smallest and the largest element of T; and finally,

$$\| \text{grad } v_1\| = 2\| z_1^T P_1\| \leqq 2\mu_M(P_1)\| z_1\| \tag{29}$$

$$\| \text{grad } v_2\| = 2\| z_2\|. \tag{30}$$

From equations (24)-(30), and considering that $\| B^* z_1\| \leqq \| B^*\| \| z_1\|$, we may now write the desired system of differential inequalities (15)

$$\dot{v} \leqq Wv,$$

where the elements of the matrix W are:

$$\begin{aligned} w_{11} &= -\mu_M^{-1}(P_1) + 2\alpha\mu_M(P_1)\mu_m^{-1}(P_1)\| B^*\| \\ w_{12} &= 2\mu_M(P_1)t_M\| z_1\| \| z_2\|^{-1} \\ w_{21} &= 2\alpha\mu_m^{-1}(P_1)\| B^*\| \| z_2\| \| z_1\|^{-1} \\ w_{22} &= -2t_m. \end{aligned} \tag{31}$$

The matrix W is stable if, and only if,

$$w_{11} + w_{22} < 0 \tag{32}$$

$$\det w > 0. \tag{33}$$

From (31) we know that

$$w_{22} < 0$$

$$w_{12}, w_{21} > 0.$$

Therefore condition (33) requires that $w_{11} < 0$ and consequently it implies (32). On the other hand, since W is a Metzler matrix condition (33),

$$\det W > 0$$

implies that W is quasidominant diagonal, i.e. that there exist positive numbers d_j, (j = 1,2), and π such that

$$d_i \, |w_{ii}| - d_j \, |w_{ij}| \geqq \pi$$

or, equivalently,

$$d_i \, |w_{ii}| - d_j \, |w_{ji}| \geqq \pi \qquad i \neq j.$$

We shall now show that the property of diagonal quasidominance of W implies stability of the overall system S [10]. Let us choose the p.d. function

$$V(z) = d_1 v_i \,(z_1) + d_2 v_2 \,(z_2) \tag{35}$$

as a Lyapunov function for the overall system S. Taking the total time derivate of V(z) along the solution of the equations (13)-(14), (and remembering that $w_{ij} < 0$ for i = j, $w_{ij} > 0$ for $i \neq j$) we have:

$$\dot{V}_1 = d_1\dot{v}_1 + d_2\dot{v}_2$$
$$\leqq d_1\ (w_{11}v_1 + w_{12}v_2) + d_2(w_{21}v_1 + w_{22}v_2)$$
$$\leqq -\ (|w_{11}| - d_1^{-1}d_2w_{21})\ d_1v_1 - (|w_{22}| - d_2^{-1}d_1w_{12})d_2v_2$$
$$\leqq -\ \tilde{\pi}\ (d_1v_1 + d_2v_2)$$
$$\leqq -\ \tilde{\pi}\ V,$$

(where $\tilde{\pi} = d_i^{-1}\ \pi > 0$).

It follows that, along the solutions of our system $V(z) > 0$ and $(-\dot{V}) > 0$, and therefore the Lyapunov "direct method" [11] guarantees that the system S is indeed stable.

7. To see how stability depends on the parameters of the model, and in particular, on the "acceleration factor" α, let us consider that $\det W > 0$, $w_{11}w_{22} > w_{12}w_{21}$. If we take the spectral norm of B^*, $\|B^*\| = (\mu_M)^{1/2}(B^{*T}B^*)$, the stability condition can be written thus:

$$t_m\{(1/2)\mu_M^{-1}(P_1) - \alpha\,\mu_M(P_1)\mu_m^{-1}(P_1)\mu_M^{1/2}(B^{*T}B^*)\}$$
$$> t_M\ \alpha\,\mu_M(P_1)\mu_m^{-1}(P_1)\mu_M^{1/2}(B^{*T}B^*)$$

or, more simply

$$\mu_M^{-1}(P_1) > 2\alpha(\mu_M)^{1/2}(B^{*T}B^*)k(P_1)(1+k(T)) \qquad (37)$$

where, given a symmetric matrix M, $k(M) \equiv \mu_M(M)\mu_m^{-1}(M)$.

Since P_1 and T are s.p.d. matrices, k can be used as a "condition number"[12].

It can readily be seen that, for sufficiently small α, i.e. for a sufficiently weak "acceleration factor', condition (36) will be verified. How small α may have to be depends on the matrices P_1, B^* and T.

For a given $Q_1 = I$, P_1 depends uniquely on H. Roughly speaking, we could take $\mu_M^{-1}(P_1)$ as an indicator of the "degree of stability" of H. Indeed if $u \in R^n$ is an eigenvector of H, we have

$$Hu = \mu\,(H)\,u \tag{38}$$

or,

$$u^T H^T = \mu(H) u^T. \tag{39}$$

Premultiplying by u^T and postmultiplying by u equation (18), remembering that $Q_1=I$ and normalizing u so that $\langle u, u\rangle = 1$, we have

$$u^T H_T P_1 u + u^T P_1 H u = -1. \tag{40}$$

Or, by making use of (38) and (39),

$$\mu(H) = -[1/2 \langle u, P_1 u\rangle] .$$

Taking into account that, $\langle u,u\rangle = 1$, we have

$$\mu_M(P_1) \geqq \langle u,P_1 u\rangle \geqq \mu_m(P_1).$$

From this we derive the following estimates for $\mu(H)$.

$$-(1/2)\mu_M^{-1}(P_1) \geqq \mu(H) \geqq -(1/2)\mu_m^{-1}(P_1)$$

so that the smaller $\mu_M(P_1)$, the smaller the upper bound of $\mu(H)$, which may be taken as some indicator of greater stability.

On the other hand, the "condition number" k can be taken as an indicator of the "degree of reactiveness" to perturbation of the relevant matrix. Therefore, the greater $k(P_1)$ and $k(T)$ $[=(t_M/t_m)]$, the less likely our system is to remain stable for a given α and a given matrix B^*.

8. It is more difficult to be conclusive concerning the <u>instability</u> conditions of the system. Intuitively, this may be understood by considering that, in order to prove stability, one tries to fix a negative upper bound for the real parts of the characteristic roots. The symmetric proof of instability would be to fix a <u>positive</u> lower bound, which is of course, redundant, and may never be true of a given system, though this is unstable.

We do know, however, that - for a sufficiently high value of the perturbation factor α, and thus, for sufficiently strong acceleration - the system (12)-(14) will lose its stability. Consider the fact that the trace of G contains elements like $\alpha b^*_{ii} \geqq 0$, where b^*_{ii} is one of the elements of the main diagonal of $B^* \geqq 0$, not all of which are zero. Those elements grow without limit as $\alpha \rightarrow +\infty$. Sooner or later the trace of G will therefore become positive, and consequently one or more characteristic roots of G will take a positive real part, leading to instability.

As we shall see below, $TrG > 0$ is an excessively strong sufficient condition for instability, which may occur for lower values of α.

9. Before discussing this point, however, we wish to emphasize some general results of the analysis so far:

(i) Since the system (as described by the matrix G) is stable for small α, and it is unstable for large α, there exists a critical value $\alpha = \alpha_c$ such that, for the matrix G, $R_e\mu_M(\alpha_c) = 0$, (μ_M being of course the root with the largest real part).

(ii) Remembering that det $G > 0$ independently of α, at $\alpha = \alpha_c$ there will be a pair of purely imaginary roots $\mu_M, \bar{\mu}_M$ (except in the very special case in which two real roots simultaneously become equal to zero).

Without loss of generality we can also assume that:

(iii) $\mathrm{Im}\mu_M(\alpha_c) \neq 0$.

(iv) The real parts of the characteristic roots other than μ_M will remain negative.

(v) $[\partial \mathrm{Re}\mu_M(\alpha)/\partial\alpha]\Big|_{\alpha=\alpha_c} \neq 0$

(i.e. the relevant roots will cross the imaginary axis with non-zero velocity).

Conditions (i)-(v) ensure that the "true" system - i.e. the nonlinear system of which the system (12) (or the equivalent (13)-(14)) is a linearization around the equilibrium point - will undergo a Hopf bifurcation, and that, besides the stationary solution z = 0, for values of α near α_c, there exists a family of periodic solutions. The amplitude of the oscillations will be of order $|\alpha - \alpha_c|^{1/2}$ and their period of order $|[2\pi/\mathrm{Im}\mu_M(\alpha_c)]|$ [13].

By analogy with what happens in the aggregate case, this means that, in a multisector multiplier-accelerator model, loss of stability leads to cyclical oscillations. Having stated this, however, we hasten to add that the proof of stability of the periodic solutions, which is already quite difficult in R^2, may be a formidable task in R^n . At any rate, it will not be attempted here [14].

10. We shall conclude our presentation with a simple numerical example in R^4, which will give the flavour of some of the problems discussed in general terms in the preceding sections. For this purpose, let us put:

$T^{(y)} = T^{(i)} = I$, (hence $G_1 = H$)

$$G_1 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix},$$

$$G_2 = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$$

$$B_0 = \begin{bmatrix} 1 & 1 \\ \\ 1 & 1 \end{bmatrix}$$

Under these assumptions, we shall have

$$G = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & -2+\alpha & \alpha \\ 0 & -1 & \alpha & -2+\alpha \end{bmatrix}$$

with the auxiliary equation

$$\det(\mu - G) = 0$$

or

$$\begin{vmatrix} \mu & 0 & -1 & 0 \\ 0 & \mu & 0 & -1 \\ 1 & 0 & \mu+2-\alpha & -\alpha \\ 0 & 1 & -\alpha & \mu+2-\alpha \end{vmatrix} = 0$$

whence

$$\mu^4 + (4-2\alpha)\mu^3 + (6-4\alpha)\mu^2 + (4-2\alpha)\mu + 1 = 0. \tag{41}$$

To find the critical value α_c, in this case we shall calculate the four Hurwitzian determinants:

$$D_1 = 4-2\alpha$$

$$D_2 = (4-2\alpha)(5-4\alpha)$$

$$D_3 = (4-2\alpha)^2(4-4\alpha)$$

$$D_4 = D_3.$$

We notice that, for α sufficiently small, $D_i > 0$, $\forall i$, and the system is stable. As we increase α continuously, two determinants, D_3 and D_4, vanish simultaneously at $\alpha=1$, whereas D_1, $D_2 > 0$. At that value of α, there will be two characteristic roots, which we shall call μ_1, and μ_2, such that

$$\mathrm{Re}\mu_1 = \mathrm{Re}\mu_2 = 0$$

$$\mathrm{Im}\mu_1 = \mathrm{Im}\mu_2 = \Omega \neq 0.$$

If we replace $\mu = j\Omega$ ($j=\sqrt{-1}$) and $\alpha = 1$ into equation (41), we have

$$\Omega^4 - j2\Omega^3 - 2\Omega^2 + j2\Omega + 1 = 0,$$

whence, equating real and imaginary parts separately,

$$\Omega^4 - 2\Omega^2 + 1 = 0$$

$$j\,2\Omega\,[1 - \Omega^2] = 0$$

or

$$\Omega = \pm 1.$$

The two purely imaginary roots are therefore

$$\mu_1 = +j; \qquad \mu_2 = -j.$$

To establish the direction of crossing of the imaginary axis, we can consider equation (41) as an implicit function

$$F(\mu,\alpha) = 0,$$

from which

$$(d\mu/d\alpha) = -[\partial F/\partial\alpha) \div (\partial F/\partial\mu)] =$$
$$[(2\mu^3 + 4\mu^2 + 2\mu)] / [4\mu^3 + 3(4-2\alpha)\mu^2 + 2(6-4\alpha)\mu + (4-2\alpha)]$$

which, at $\mu = +j$ and $\alpha = 1$, gives

$(d\mu/d\alpha) = 1 > 0.$

(We may get an analogous result by putting $\mu = -j$).

The crossing will therefore take place with non-zero velocity. We may then conclude that the conditions for Hopf bifurcation, listed under § 9, are verified here.

11. CONCLUSION

Initially we presented an aggregate multiplier-accelerator model and showed that, when stability is lost due to a strong accelerator, oscillations will ensue. Next, a multisector version of the model was developed. Stability conditions were analyzed by means of the vector Lyapunov functions method. In particular, the effect on stability of changing parameters was investigated. Finally we showed that, for critical values of the parameters, the system undergoes a Hopf bifurcation and has self-excited periodic solutions. A major result of the aggregate trade cycle theory of the Keynesian type has therefore been generalized to the n-dimensional case.

NOTES

[1] An interesting discussion of the priority of this result in bifurcation theory can be found in Arnold [1983], pp. 271-273. In this paper, however, we shall make use of the more common denomination of "Hopf bifurcation". On this subject cf. Marsden and Mc.Cracken [1976] and Hassard, Kazarinoff and Wan [1981] .

[2] Cf. Marsden-Mc.Cracken, op.cit., pp. 63-83
Hassard-Kazarinoff-Wan, op.cit., pp. 14-24

[3] For an extensive discussion of this point, cf Medio [1979] , Chapters II and V.

[4] Cf. Gantmacher, [1966], vol 1, pp. 46-47.

[5] We assume, in particular, that $b_{ii} > 0$ for some $i = 1, 2,...,n$.

[6] On the concept of D-stable matrices and related issues, see Magnani-Meriggi [1981] , pp. 535-544.

[7] An excellent and thorough discussion of this method can be found in Siljak [1978] , on which this author has drawn heavily.

[8] Cf. Gantmacher, op.cit., vol. 2, pp. 182-183.

[9] Uniqueness of solution requires that no characteristic root is equal to zero, and that for no pair of characteristic values of μ_i and μ_j we have $\mu_i = -\mu_j$, (which we assume here).

[10] Cf. Siljak, op.cit., pp. 39-40 and 96-99.

[11] On the Lyapunov direct method, cf La Salle and Lefschetz [1961] .

[12] On the concept of "condition numbers", see, for example, Stewart [1973], pp. 184-192.

[13] A more rigorous statement of the conditions of the Hopf bifurcation may be found in Marsden-Mc. Cracken and Hassard-Kazarinoff-Wan, loc.cit..

[14] The interested reader may consult Marsden-Mc. Cracken, cit., pp. 104-131 on this point, where an algorithm is provided for the determination of stability of periodic solutions. Unfortunately, in the general case, this involves estimating the sign of derivates up to at least third order, which may not be possible in an economic model on purely a priori grounds.

REFERENCES

Arnold, V.I., Geometrical Methods in the Theory of Ordinary Differential Equations, Springer-Verlag, New York, 1983.

Gantmacher, Théori des matrices, Dunod, Paris, 1966.

Hassard, B.D., Kazarinoff, N.D., Wan, Y.H., Theory and Applications of Hopf Bifurcation, Cambridge University Press, 1981.

La Salle, J., Lefschetz, S., Stability by Lyapunov's Direct Method, with Applications, Academic Press, New York, 1961.

Magnani, U., Meriggi, M.R., "Characterization of K-Matrices", in Castellani, G, Mazzoleni, P. (eds.), Mathematical Programming and its Economic Applications, Milano, 1981.

Marsden, J.E., Mc. Cracken, M., The Hopf Bifurcation and Its Applications, Applied Mathematical Sciences, vol. 19, Springer-Verlag, 1976.

Medio, A., Teoria nonlineare del ciclo economico, Il Mulino, Bologna, 1979. (The Italian version of an unpublished Ph.D. thesis, Cambridge, U.K., 1975).

Siljak, D.D., Large-Scale Dynamic Systems, North-Holland, Amsterdam, 1978.

Stewart, G.W., Introduction to Matrix Computations, Academic Press, New York, 1973.

Chapter 13

(M,R)-Systems as a Framework for Modeling Structural Change in a Global Industry

J. CASTI

1. THE PROBLEM OF STRUCTURAL CHANGE

With the arrival of rapid global communication facilities via satellites, and the widespread availability of cheap worldwide transportation by air and sea, the phenomenon of the transnational corporation (TNC) has emerged, carrying with it a total and complete re-shaping of the structure and operation of major industrial activities. Prior to the advent of the TNC, large industries were considered nationally; now, for example, the automotive industry is scattered throughout the world with firms engaging in design in one place, production in another, and marketing and sales everywhere. The situation is further compounded and confused by a myriad of interlocking joint ventures, co-production agreements, partial mergers and so forth. What all of this amounts to is a discontinuous shift from one way of doing business to another, and from one industrial paradigm, emphasizing national centralization and domestic markets, to a global structure transcending national boundaries and, to a great extent, local governmental control. The problem of industrial structural change is basically how to account for this transition, how to understand its implications for the future evolution and development of a given industry, and how to gain some understanding of the way such changes can be directed, or managed, to avoid unnecessary chaos, disorder and economic upheavals during the transition periods from one structure to another.

To study the problem of structural change, a suitable conceptual framework is needed, within which the various firms comprising the industry can play out their roles in interaction with their environment. Whatever framework is used, it must account for the way in which firms execute their design, production and marketing functions, and incorporate mechanisms whereby the firms can expand, merge, or even cease to exist. The conceptual scheme must also allow for the mutual interactions of the firms of the industry, both with each other, and with their external environment. The outside environment includes the suppliers of the raw materials and resourses needed for the firm's activities, the consumers of the firm's product and the various environmental influences exerted by government regulators acting by the setting of the economic climate (through taxes, interest rates, exchange restrictions, etc.), and the business climate (through tariffs, quotas, import restrictions and the like).

The foregoing requirements for a conceptual framework for the study of industrial structural change have a strongly biological overtone, suggesting that the view of a global industry as a living multicelled organism may serve as a foundational metaphor for the framework we seek. The balance of this paper is devoted to the exploration of this idea. More specifically, the notion of an (M,R)-system (metabolism-repair system) is examined as a candidate for the type of theoretical construct needed to capture the main features of industrial structural change. Originally, (M,R)-systems were introduced into biology by Rosen (1958a, 1958b) as a means to study cellular development of organisms by breaking away from the traditional bio-chemical types of analyses, and employing a purely relational analysis emphasizing the functional rather than structural organization of the system. This approach leads to the study of <u>classes</u> of abstract biologies and a means for their comparison rather than to detailed material analysis of a single organism. This is exactly the type of scheme needed to investigate industrial structural change, although as we go along it will become clear that certain biological aspects of the (M,R)-systems will require modifications in the industrial context. Nonetheless, the (M,R)-framework that follows, does, in our opinion, provide a suitable mathematical skeleton upon which to build an operational theory of industrial structural change.

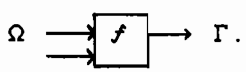
2. PRODUCTION AND SALES AS AN INPUT/OUTPUT PROCESS

Our underlying basic hypothesis is that an industry such as the world auto industry, is composed of a collection of interacting firms receiving inputs from an external environment, processing these inputs into the products of the firm, which are in turn discharged back into the environment for which the firms receive additional resources (money, usually) to continue their activity. For a variety of reasons which will become clear later, it is most natural and convenient to regard the firms' outputs as the money they receive from their products rather than the products, themselves. In short, the real mission of a firm is to make money, not products, and the products are thought of as only a vehicle to facilitate this higher-level goal.

For the moment, let us concentrate upon the description of a single firm F as an (M,R)-system. We will return later to the case of several firms (an industry). Let Ω denote the set of environmental inputs received by F . In general, the elements of Ω are both *physical* inputs - such as raw materials, labor, machinery and so forth, and the external *operating* inputs - such as the economic, political and technological constraints of the general environment. The firm accepts an input $\omega \in \Omega$ and processes it via some internal production and marketing procedures to produce a marketable product which is then sold, thereby generating an output $\gamma \in \Gamma$, measured in monetary units. Here, due to the assumption that the observed output is money, we could take $\Gamma = R$, the real numbers. To maintain uniformity of scale between inputs and outputs, we can introduce prices for all environmental inputs, thereby converting all inputs into monetary equivalents. We shall omit this mathematically trivial, but possibly economically important step in the remainder of the paper. Thus, abstractly, the behaviour of F is represented by a *metabolic* map

$$f : \Omega \rightarrow \Gamma .$$

(Note: in the economics literature, it is common to call f a *production function*. To preserve our biological metaphor, we shall depart from this convention in this paper). Schematically, we can represent the production and sales component of F as



The foregoing picture is very familiar in the mathematical system theory literature, where it is termed an *external* or *input/output* description of the behavior of F (see Casti 1987, Kalman 1969). If we want to focus attention on the manner in which the inputs from Ω are translated into revenue by means of specific products of the firm, then we must look at the *internal* behavior of F . Abstractly, what this means is that we must "factor" the metabolic map f through a state space X , using two maps g and h

such that

$$g : \Omega \rightarrow X, \quad h : X \rightarrow \Gamma .$$

In other words, we must find a space X and maps g and h such that the diagram

$$\begin{array}{ccc} \Omega & \xrightarrow{f} & \Gamma \\ & {}_{g}\searrow \quad \nearrow_{h} & \\ & X & \end{array}$$

commutes.

In our industrial setting, the space X and the maps g and h have a very interesting interpretation: X represents the actual products which the firm produces (cars, TVs, lamps, drugs or whatever), while the map g specifies how inputs are transformed into products (a *production* map). The map h represents the manner in which the firm translates products into revenue, (i.e., h is a *marketing/sales* map).

For a variety of practical as well as mathematical reasons, it is customary to impose the additional requirements that the map g be *onto*, while the map h is *one-to-one*. Such a factorization of f is called *canonical*, and is essentially unique. These conditions have a very direct interpretation in the business setting: the production map being onto, means that any level of production can be achieved if F is supplied with suitable inputs, i.e. there are no intrinsic limitations on the firm's ability to produce products given adequate raw materials and other resources. The requirement that the marketing/sales map be one-to-one just means that different levels of production generate different amounts of revenue. Or, put another way, two distinct levels of production cannot generate the same revenue for F . Producers for generating such canonical factorizations of a metabolic map f are well-known in the system theory literature and will not be discussed further here (cf. Casti 1987, 1985; Kalman 1969).

3. INNOVATION, REPAIR AND REPLICATION

The standard system-theoretic framework presented above for describing the metabolic behavior of a firm as an input/output map f (or, by its equivalent factorization through the production and marketing/sales maps g and h) would be perfectly adequate for the characterization of F if the firm were operating in a totally stable environment with no competition. However, the intrusion of real-world considerations into the firm's activities results in the need for the firm to continually engage in changes of its product, introduction of new production techniques and development of alternate marketing strategies if it is to remain a viable enterprise. In biological terms, the firm must repair damages and adapt or else enter a senescent phase ultimately resulting in its extinction. In biological organisms, the adaption and repair is carried out by genetic programs which re-process the system output to renew the metabolic behaviour of the organism. In the context of an auto firm, such a restoration of the metabolic activity can only result from the repair of different production and/or marketing procedures, i.e. renewal of the maps g and h. This can come about only through technological improvements, better managerial procedures and/or incorporation of new knowledge, i.e. through innovation. Our basic question here is: how can the modeling framework introduced earlier be extended in a natural fashion to account for the firm's need to "innovate or die?"

The key to answering this question is to note that the only way the firm can renew its metabolic activity, is to utilize some part of its revenues to regenerate either its production processes, or its marketing approach, or both. Thus, the firm's "repair" mechanism must ultimately be a map which transforms the firm's output (revenues) into the desired metabolic structure. If we let $H(\Omega,\Gamma)$ denote the set of all possible metabolic processes, and if Φ_f denotes the repair map, we have

$$\Phi_f : \Gamma \longrightarrow H(\Omega,\Gamma).$$

Note that we explicitly indicate the dependence of the repair map upon the metabolism f, since the objective of the repair procedure is to reproduce f which is an activity of the firm and, as such, is affected by the metabolic activity. We now have the following abstract diagram for a firm F as a metabolism-repair (M,R)-system.

$$\Omega \xrightarrow{f} \Gamma \xrightarrow{\Phi_f} H(\Omega,\Gamma).$$

As already noted, f represents the firm's procedures for operating upon the environment to produce or "metabolize" revenues, while Φ_f corresponds to the firm's "genetic" capacity to repair disturbances in metabolism arising from environmental fluctuations.

To complete the metaphor of the firm as a biological organism, we must address the issue of how to repair the repairers. The repair mechanisms were introduced to account for the fact that during the course of time, the firm's metabolic machinery will erode and decay, thereby requiring some sort of rejuvenation if the firm is to avoid extinction. Precisely the same argument applies to the repair mechanism, but it is of no particular help to introduce repairers for the repairers and so forth, ending up in a useless infinite regress. The way out of this loop is to make the repair components self-replicating. In this way, new copies of the repair mechanism are continually being produced, and it is unnecessary to assume the repair functions are immortal or to fall into an infinite regress of repairers to insure survivability of the firm. Let us see how to introduce the idea of replication into the foregoing framework.

Since the replication operation involves reproducing the genetic component Φ_f from the metabolic activity of the firm, it follows that the replication map, call it β_γ must be such that

$$\beta_\gamma : H(\Omega,\Gamma) \rightarrow H(\Gamma, H(\Omega,\Gamma)),$$

if it exists at all. The question is: how can such a map β_γ be constructed from the basic metabolic components Ω ,Γ, and $H(\Omega,\Gamma)$ of the firm? To see how this is done, it is easiest to consider a somewhat more general situation.

Let X and Y be arbitrary sets. Then for each $x \in X$, we can define a map

$$\hat{x} : H(X, Y) \rightarrow Y$$

by the rule

$$\hat{x}(f) = f(x),$$

for all $f \in H(X, Y)$. Thus, we have an embedding of X into the set $H(H(X, Y)\, Y)$. Now, assume that the map $\hat{x}$ has a left inverse $\hat{x}^{-1}$, so that

$$\hat{x}^{-1} : Y \rightarrow H(X, Y).$$

Then, we clearly have

$$\hat{x}^{-1}\, \hat{x}\,(f) = f$$

for all $f \in H(X, Y)$.

Returning now to our replication situation, we set

$$X = \Gamma, \quad Y = H(\Omega, \Gamma),$$

and apply the foregoing general argument to obtain for each $\gamma \in \Gamma$, a map $\beta_{\gamma} = \hat{\gamma}^{-1}$ with the property that

$$\beta_{\gamma} : H(\Omega, \Gamma) \rightarrow H(\Gamma, H(\Omega, \Gamma))$$

for all $\hat{\gamma}$ possessing a left inverse. In short, the metabolic activity of the firm can be used to reproduce its repair component, if the technical condition on the invertibility of the map $\hat{\gamma}$ is satisfied. The economic interpretation of this condition is that $\hat{\gamma}$ is invertible if different innovations and R&D activities (i.e. different genetics mechanisms $\Phi_{f_1} \neq \Phi_{f_2}$) give rise to different production and marketing functions(i.e. different metabolic processes $f_1 \neq f_2$). In the industrial context we are examining, this seems to be a reasonably defensible assumption which will be accepted for the remainder of the paper.

Before entering into a more thorough discussion of the implications of the (M,R)-system as a paradigm for industrial structuring and operation, it is of interest to consider the actual meaning of the replication process described by the map β_γ. We have seen that the repair mechanism Φ_f basically provides the prescription by which revenues are used to support and renew the production-marketing process f. By the same token, the replication process β_γ gives the instructions by which the genetic process Φ_f is duplicated. Thus, since Φ_f corresponds to innovation/R&D, we can only conclude that β_γ corresponds to the diffusion of innovation/R&D. In short, β_γ is a prescription for growth of the firm by development of new divisions. Alternately, it could represent the start-up procedure of new firms that spin-off from the parent corporation. In either case, the innovation and "know-how" of the parent firm is transferred to a new organization, and is then used as part of the metabolic operation f to produce revenue from environmental inputs in the usual way.

4. CONSEQUENCES OF THE (M,R)-FRAMEWORK FOR A SINGLE FIRM

The minimal structure introduced thus far to define an (M,R)-system is already sufficient to shed light on a variety of interesting questions surrounding the way a firm can respond to changes in its operating environment, the possibility for innovation to occur through environmental effect, the circumstances under which environmental changes can be reversed, feedback, and so on. In this section we sketch the way in which these issues appear within the (M,R)-framework, and consider the conclusions that can be drawn about firm behavior from this structure.

A. Stable Metabolic Operations in Changing Environments - imagine the situation in which the firm's "usual" input ω of raw materials, labor, etc. is disturbed to a new input $\bar{\omega}$. The condition for stable operation of the firm is for the environment ω to be such that

$$\Phi_f\,(f(\omega)) = f \quad , \qquad (*)$$

i.e. the metabolic structure f is stable in the environment ω in the sense that the repair mechanism Φ_f always regenerates f when the environmental input is ω. We would say that all $\omega \in \Omega$ satisfying (*) form a stable environment for the firm.

Now suppose that the new environment $\bar{\omega} \neq \omega$. Then (*) will only hold if either

$$f(\omega) = f(\bar{\omega}) \quad \text{or} \quad \Phi_f(f(\bar{\omega})) = f.$$

The first case is trivial in the sense that the observed revenues of the firm are invariant to the change of environmental inputs. If $f(\omega) \neq f(\bar{\omega})$ then the firm's revenues are not stable with respect to the change of environment and we must consider the repair mechanism to see whether or not the environmental alterations can be compensated for in the sense that

$$\Phi_f\,(f(\bar{\omega})) = \bar{f} \neq f,$$

with $\bar{f}(\bar{\omega}) = f(\omega)$, i.e. will the genetic mechanism produce a new metabolism $\bar{f}$ which will duplicate the revenues of f, but with the input $\bar{\omega}$ rather than ω? In this case, the entire metabolic activity of the firm would be permanently altered if we had

$$\Phi_f(\bar{f}(\bar{\omega})) = \bar{f}.$$

On the other hand, if we had $\bar{f}(\bar{\omega}) = f(\omega)$ or, more generally,

$$\Phi_f(\bar{f}(\bar{\omega})) = f,$$

then the firm's metabolism would only undergo periodic changes in time.

Finally, we could have the situation in which

$$\Phi_f(\bar{f}(\bar{\omega})) = \hat{f} \neq f, \bar{f}$$

and, iterating this process, we see that an environmental change may cause the firm to wander about in the set $H\,(\Omega, \Gamma)$, changing its production-marketing procedures through a sequence of metabolic processes $f^{(1)}, f^{(2)}, f^{(3)}, \ldots$. This "hunting" process will terminate if either

(i) there exists an N such that
$\Phi_f(f^{(N)}(\bar{\omega})) = f^{(N)}$

or

(ii) there exists an N such that
$\Phi_f(f^{(N)}(\bar{\omega})) = f^{(N-k)}$, $k = 1,2,\ldots,N-1.$

In case (i) the firm becomes stable in the new environment $\bar{\omega}$, while in case (ii) the firm undergoes periodic changes in its metabolic structure. If no such N exists, the firm is unstable and aperiodic. (Note: This last possibility can occur only if the set $H(\Omega,\Gamma)$ of possible production-marketing procedures is infinite.

B. "Lamarckian" Changes in the Repair Process – the above discussion of metabolic change was undertaken subject to the tacit assumption that the repair map Φ_f remains unchanged. It is of interest to inquire as to whether or not an environmental change $\omega \rightarrow \bar{\omega}$ can generate a "Lamarckian" type of genetic change in Φ_f through the replication process described in the last section. If such a change were indeed possible, then it would imply that the actual innovation/R&D process, which regenerates the metabolic activity f, could be affected by environmental changes alone.

To examine this question, suppose we have the environmental change $\omega \rightarrow \bar{\omega}$. Then the replication map β_γ associated with the input ω and the output $\gamma = f(\omega)$ is changed to $\beta_{\bar{\gamma}}$ where $\bar{\gamma} = f(\bar{\omega})$. Recalling that

$$\beta_\gamma = \hat{\gamma}^{-1}, \quad \beta_{\hat{\gamma}} = \hat{\gamma}^{-1},$$

and

$$\hat{\gamma}(\Phi_f) = \Phi_f(f(\omega)), \quad \hat{\gamma}(\Phi_f) = \Phi_f(f(\bar{\omega})),$$

after applying $\beta_\gamma, \beta_{\hat{\gamma}}$, respectively to the last two relations, we find that

$$\beta_{\gamma}(\Phi_f(f(\omega))) = \beta_{\bar{\gamma}}(\Phi_f(f(\bar{\omega}))) = \Phi_f,$$

showing that the new replication map $\beta_{\bar{\gamma}}$ replicates the existing repair component Φ_f exactly. Thus, an environmental change alone can have no effect upon the repair map Φ_f.

Now we ask whether it is possible for a change in the metabolic production-marketing procedures to result in a change of the firm's "genotype". Suppose we replace the metabolic activity f by some other production-marketing process $b \in H(\Omega,\Gamma)$. By definition

$$\hat{b}(\omega)(\Phi_f) = \Phi_b(h(\omega)).$$

Assuming $\hat{b}(\omega)$ is invertible, we apply $\hat{b}^{-1}(\omega)$ to both sides of the above relation to obtain

$$\beta_{b(\omega)}(\Phi_f(b(\omega))) = \Phi_f.$$

Thus, the induced replication map reproduces the original repair component of the firm under all conditions. In short, no Lamarckian changes in the metabolic component, either in the environment Ω or in the metabolic set $H(\Omega,\Gamma)$, can result in changes in the firm's repair mechanism. Such "genetic" changes can only come about through a direct intervention in the genetic code itself (mutation), and not via indirect metabolic alterations.

C. Feedback as an Environmental Regulator – the environmental changes discussed thus far have been assumed to be generated by actions external to the firm; however, it may often be the case that the firm's output of revenue is employed as one of the environmental inputs, i.e. ω is a function of γ, $\omega = \omega(\gamma)$. In this event, the firm actually creates part of its own environment and, as a consequence, can partially regulate its own structural alterations. An important aspect of this general process is to understand as to what degree adverse environmental disturbances can be "neutralized" by suitably chosen feedback policies. This question is a special case of the more general problem of "reachability", in which we ask about the possibility of attaining any pre-defined metabolic structure by means of a sequence of environmental changes. This is a topic we shall return to later in connection with discussing the dynamical aspects of the firm's morphology.

5. GLOBAL INDUSTRY AS A NETWORK OF (M,R)-SYSTEMS

So far we have only considered a single firm F as an (M,R)-system. If we connect several firms together, with the outputs of some firms serving as inputs to others, and the repair mechanisms of each firm requiring the output from at least one firm, then we have the structural basis for characterization of an entire industry. Such a network of interdependent firms gives rise to a number of significant questions involving the birth, growth and death of an industry and of the individual firms comprising the industry. In this section, we consider how these issues arise naturally within the context of an (M,R)-network, and the way in which our earlier (M,R)-formalism for a single firm can be extended to form a basis for modeling an entire industry.

In order to fix ideas, consider the specific (M,R)-network depicted in Figure 1. The square blocks labeled $F_1, F_2, ..., F_6$ represent the metabolic processes of the individual firms, while the ovals, denoted $R_1, R_2, ..., R_6$, represent the respective firms' repair mechanisms. The requirements that we impose for any such network are modest:

i) each firm must receive at least one input, either from the external world or from the output of another firm;
ii) each firm produces at least one output;
iii) each repair mechanism receives the output of at least one firm in the network.

We have stated the requirements for an (M,R)-network in quite general terms. In a typical industry, such as the automotive industry, it is likely that each firm will receive its input from the outside world and discharge at least part of its output of revenues back to the external world in the form of payment for goods and services and returns to shareholders. Also, the repair mechanisms will, for the most part, receive only the output from their corresponding firm, since the case of one firm devoting its resources toward supporting another (as with the firm F_4 of Figure 1)

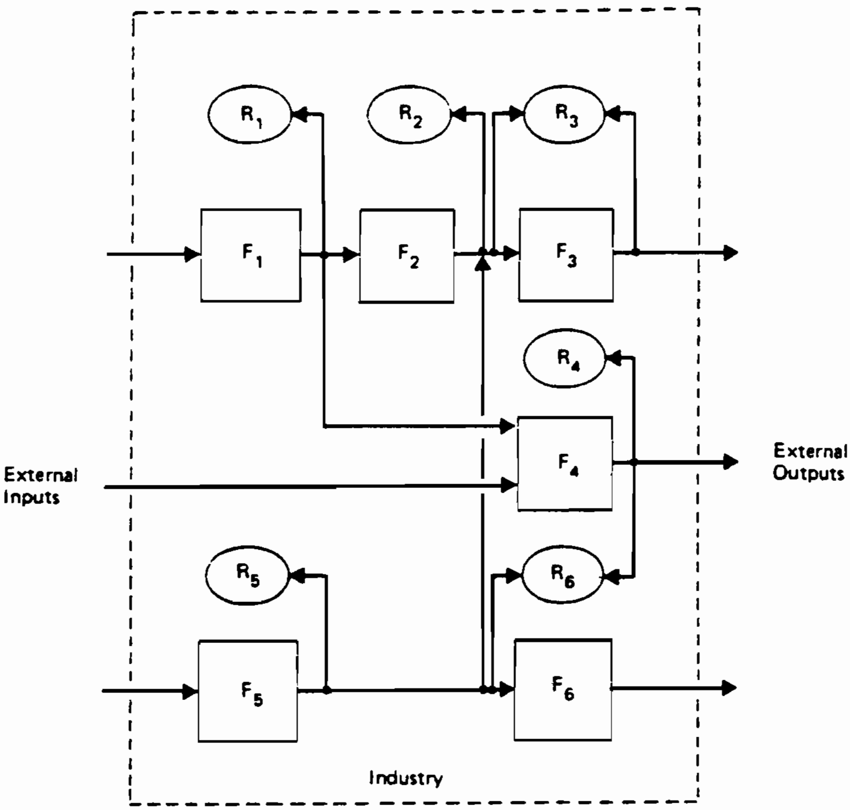


Figure 1 A Typical (M,R)-Network

seems rather improbable under most circumstances, although certainly not impossible.

The first general issue to consider for an (M,R)-network is the dependency structure. We are concerned with the question of how the removal of a given firm from the network affects the existence and operation of other firms in the industry. For instance, referring to Figure 1 we see that the failure of F_5 results in the failure of firm F_6 , as well, since F_6 receives its only input from F_5. Furthermore, the failure of F_5 may influence the operation of F_3 , as F_3 receives part of its input from F_5 . Thus, in this case we would consider firm F_3 and F_6 to comprise the dependency set of firm F_5. Any firm whose failure affects either the existence or operation of all firms in the industry will be called a central firm, i.e. the dependency set of a central firm is the entire industry.

Since we know that in the absence of the repair mechanism any firm will go out of existence after some finite lifetime, it is clear that other firms in the dependency set of a given firm will also go out of existence when that firm does. However, with the repair mechanism in operation, it is quite possible that a given firm could "come back to life" even after its initial demise. For example, firm F_6 in Figure 1 may cease metabolic operation and be removed from the network; however, the repair mechanism R_6 receives its necessary inputs from firms F_4 and F_5 indicating that whatever "shock" caused the extinction of F_6, the firm will be re-inserted into the network after some characteristic delay time depending upon the repair mechanism R_6. In other words, copies of F_6 will continue to be manufactured even after the removal of F_6 from the network. Firms like F_6 will be called re-establishable, while all other firms are termed non-reestablishable (e.g. F_1, F_5, ...). There is an important relationship between the notion of re-establishability and the concept of a central component expressed by the following result.

Theorem 1. Every (M,R)-network must contain at least one non-reestablishable firm.

Corollary. If an (M,R)-network contains only one non-reestablishable firm, then that firm is a central component. The proofs of these results can be found in the papers cited in the references.

The significance of this result is twofold:

i) every industry must contain at least one firm whose metabolic failure cannot be repaired. This conclusion follows only from the connective structure of the (M,R)-network and is completely independent of the specific industry, the procedures of the firms, their products or marketing strategies. It is solely a consequence of the meaning of the metabolism-repair functions and the replication process.

ii) in order to be "resilient" to unforeseen disturbances, one would desire an industry to consist of a large number of re-establishable firms. On the other hand, the above results show that if only a small number of firms are non-reestablishable, then there is a high likelihood that one of them will be a central component whose failure will destroy the entire industry. Thus, an industry with a large number of re-establishable firms will be able to survive many types of shocks and surprises, but there will be certain types of disturbances which will effectively cripple the whole industry. Consequently, it may be better to have an industry with a relatively large number of non-reestablishable firms if it is desirable to protect the industry from complete breakdown.

6. TIME-LAGS AND DYNAMICS

Up to this point, it has been assumed that the metabolic and repair functions of the firm take place instantaneously, i.e. inputs are transformed into revenues immediately, and there is no delay in either repairing the metabolic process itself or in the replication of the repair mechanism. Needless to say, these assumptions are pure fiction; production of revenue and repair/replication takes time and the delays involved often spell the difference between success or failure for a firm.

While there is no space here to enter into a detailed discussion of the matter, let us simplify the situation by assuming only two types of delays. The first we term the production delay, corresponding to the time required to transform a given input of materials, manpower and knowledge into an observable amount of revenue, or the time required for a repair function to restore a metabolic operation. The second type of delay we shall call the transport delay. It corresponds to the time needed to transport the output of a firm to where it can be utilized as the input to either another firm or a repair mechanism (or to the external world).

As an illustration of how time-lags can influence the behavior of an (M,R)-model of a firm, consider the case of the single firm F depicted in Figure 2. The firm is clearly non-reestablishable in the sense discussed earlier. If the combined delay time of the

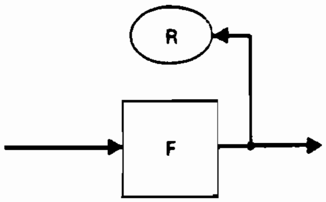


Figure 2 A Single Firm

production delay of R and the transport delay from F to R is T units, and if at time $t=0$, F produces an output and is then removed from the network, T units later R will produce a copy of F and F will be built back into the network even though F is graph-theoretically non-reestablishable. However, if F is not just removed from the industrial network, but is suppressed for an amount of time $t \geqq T$, then irreversible damage will have occurred and F will be removed from the network forever. The interplay between the various time-lags involved when several firms are coupled together into an industry is a delicate matter and will be taken up in a later paper.

Closely related to the time-lag problem is the matter of system dynamics. There are several issues surrounding this topic, not all of them mutually consistent. For simplicity, let us consider here only the case of a single firm F modeled as an (M,R)-system. Abstractly, the diagram for F is

$$\begin{array}{ccccccc} \Omega & \xrightarrow{f} & \Gamma & \xrightarrow{\Phi_f} & H(\Omega,\Gamma) & \xrightarrow{\beta_y} & H(\Gamma, H(\Omega,\Gamma)) \\ & g\searrow & & \nearrow h & & & \\ & & X & & & & \end{array}$$

and ask in what manner F can be regarded as a dynamical system. If it were not for the repair and replication maps Φ_f *and* β_y this would be a straightforward question addressable via normal system-theoretic realization theory procedures; i.e. we would have the problem of constructing a canonical internal model of the firm

$$\dot{x} = p(x, u)$$
$$y = h(x),$$

whose input/output behavior duplicates that of the given metabolic map f. Techniques for handling this question are readily available in the mathematical system theory literature (Casti 1987, 1985; Kalman 1969).

Let us ignore for the moment the factorization of the firm's metabolism f through the production-marketing maps g and h, and consider the (M,R)-system

$$\Omega \xrightarrow{f} \Gamma \xrightarrow{\Phi_f} H(\Omega,\Gamma)$$

where $f \in H(\Omega,\Gamma), \Phi_f \in H(\Gamma,H(\Omega,\Gamma))$. We wish to show how this abstract model of a firm can be considered as a sequential machine, i.e. as a discrete-time dynamical input/output system.

Let us recall that a sequential machine M is a composite $M = (A, B, S, \delta, \lambda)$, where A, B, and S are sets (possibly infinite), while $\delta : A \times S \to S$, $\lambda : S \times A \to B$ are maps. We interpret A as the input alphabet of M, B as the output set, S as the set of states, with δ and λ the state- transitions and output maps of the machine, respectively. At each discrete instant of time $t = 0,1,2,...,$ M receives an input symbol from A, emits an output in B, and the state is changed according to the rule δ, and the process continues from the time $t + 1$. Further details on the properties of sequential machines can be found for example in (Kalman 1969; Eilenberg 1974).

In order to characterize the firm F as a sequential machine, we make the identifications

$$A = \Omega, \quad B = \Gamma, \quad S = H(\Omega,\Gamma).$$

$$\delta(\omega, f) = \Phi_f(f(\omega)), \quad \lambda(f, \omega) = f(\omega).$$

Thus, in general any firm can be regarded as a sequential machine in which the set of "states" of the machine correspond to the set of possible "phenotypes" of the firm, while the input and output sets of the machine are the inputs and revenues of the firm, respectively.

Putting the above ideas together, we arrive at the following scheme for characterizing the dynamics of the firm:

i) regarding the firm as a sequential machine formed from the elements $F = (\Omega, \Gamma, H(\Omega, \Gamma), \Phi_f, f)$, we compute the metabolic process f at the time $t = 0$;

ii) using f, Ω, Γ, we employ realization theory to form a canonical model for the state space X and the production/marketing maps g and h;

iii) Let $t \to t + 1$ and use the sequential machine to calculate the new metabolism f. It is the same as at the previous time-step. Then continue to use the earlier X, g and h; if f changes, calculate a new canonical production/marketing model and continue the process with the new model until the next time period.

It is of interest to note that in the above scheme, a change of metabolism implies that the production process, the marketing procedure and/or the actual product has been changed. This can come about only if the repair map Φ_f fails to reproduce f. We have already seen that this may come about only by means of environmental changes, in general, unless the replication map fails to exist. But this last situation depends entirely upon the size of the set $H(\Omega, H(\Omega, \Gamma))$, the space of all possible repair maps. If it is either too large or too small, then no replication is possible. It would take us too far afield to enter into the details of this argument here, but the implications are that it is only in a highly restricted class of categories that replicating (M,R)-systems exist, and it is within this class that we must search for viable models of industrial growth and decline.

7. ATTAINABLE PRODUCTION-MARKETING PROCESSES

The arguments given earlier show that the metabolic component of a firm can be altered by changes in its environmental inputs, while such changes leave the firm's repair mechanism unaffected. By turning these arguments around, we can investigate the degree to which environmental changes can be used in order to bring the firm's production and marketing processes to some pre-assigned state. An important special case of this "reachability" question is to ask if a metabolic structure reached by some sequence of environmental changes can be reversed by another appropriate sequence of changes in the environment. Questions of the above type strike to the heart of many important industrial issues having to do with the way in which changes in materials, men, and machines can be employed to affect the overall productive capabilities of the firm.

In terms of machines, the reachability problem can be stated as: given a machine in a specified initial state, does there exist a sequence of inputs which will bring the machine to some preassigned state (perhaps also at a preassigned time)? In general, the answer to this question is no. Machines having this "complete reachability" property are called strongly connected, and we can ask whether or not machines which correspond to (M,R)-systems are strongly connected.

Generally speaking, machines corresponding to (M,R)-systems may fail to be strongly connected; hence, there may exist abstract "firms" which may be unreachable from any initial configuration by any sequence of environmental alterations. In the usual theory of sequential machines, this difficulty can be formally by-passed by enlarging the set of inputs and by appropriately extending the maps δ and λ. For (M,R)-systems this is a much more subtle business for the following reasons:

(a) the input set Ω and the state set $S = H(\Omega, \Gamma)$ are related in the (M,R)-systems and we cannot enlarge Ω without also enlarging S;

(b) by extending the maps Φ_f and f in the (M,R)-systems, we move the mappings from the sets $H(\Omega, \Gamma)$ and $H(\Gamma, H(\Omega, \Gamma))$ to new sets $H(\Omega', \Gamma), H(\Gamma, H(\Omega', \Gamma))$, respectively. But this last set must possess certain properties in order for replication to be possible and this property is by no means implied by the replicability of the original system.

8. PROSPECTS AND CONCLUSIONS

The development of (M,R)-systems as a theoretical framework for the study of industrial growth and re-structuring has only been tentatively sketched in the preceding pages to the degree necessary to demonstrate feasibility of the idea. To transform the basic idea into a working tool to study, for instance, the evolution and development of the world automotive industry, requires a substantial research effort on both the theoretical, as well as applied fronts. It will be necessary to give concrete meaning and structure to the various abstract components composing the (M.R)-network (the elements Ω, Γ, $H(\Omega, \Gamma)$, Φ_f, etc.), as well as working out the various connectivity structures which link the individual firms comprising an industry. Such activities form the basis of the applied component of any implementation of the (M,R)-framework for a specific industry. Some complementary evolutionary ideas are given for the auto industry by Businaro (1982) and their connection with (M,R)-systems merit much further study. See also the general evolutionary ideas in Nelson and Winter (1982) and Businaro (1983).

But there are also a number of purely theoretical aspects of the (M,R)-formalism which need further study if the overall structure is to bear the weight of providing the foundation for such an investigation of individual dynamics. We have already touched upon some of these issues in passing, but it is worthwhile to re-examine them again as the basis for a future research agenda.

i) <u>Lamarckian Changes</u> - we have seen that changes in the firm's repair mechanism Φ_f cannot come about by environmental alterations alone, as long as certain invertibility assumptions on the replication procedure hold. This assumption, and its resultant conclusion, are quite acceptable in the biological context but rest upon much shakier ground in our industrial setting. It certainly seems plausible that at least certain types of environmental changes could give rise to a change of the firm's "genotype". At this stage it is unclear exactly how to modify the mathematical setting given above to accommodate such "Lamarckian" changes.

ii) <u>Networks and Time-Lags</u> - the manner in which time-lags, in both firm operations and in transport from one firm to another, affect the overall behavior of an industry is critical for determination of the long-term growth or decline of given firms within the industry. We have already seen simple examples in which time-lags can result in either the permanent extinction of a firm or, conversely, in its "resurrection" after being theoretically "dead". The interdependencies of lags of different types and lengths is a topic that cannot be ignored if the (M,R)-framework is to be used to gain insight into the behavior of real industries.

iii) <u>Dynamics</u> - the procedure outlined in the text for regarding an (M,R)-system as a sequential machine is one way to introduce dynamical considerations into the overall formalism. There may be many other non-equivalent approaches, each leading to a different view of the dynamical behavior of a firm. Even accepting the approach given here for a single firm, there still arises the question of what will be the dynamical behavior of a collection of such firms, i.e. an industry. Obviously, the answer to such a question depends upon the connective and dependency structure of the network, which in turn takes us back to some of the time-lag considerations discussed earlier.

iv) Adaptation and Selection - if an (M,R)-network is to provide a mathematical metaphor for the evolution of an industry, then it must possess some means to accommodate the concepts of genetic variability and adaptive selections. We have already spoken of the need to be able to incorporate genetic changes in the repair map Φ_f into the mathematical machinery of (M,R)-systems. A natural candidate for the selection mechanisms is to impose some sort of optimality criterion upon the possible abstract firms which may result from genetic "mutations". Production efficiency, profitability, survivability are logical possibilities, but so also are less economic-oriented criteria like degree of re-establishability and level of centrality, criteria suggested more by the functional role of a firm in a network than by its economic performance as an isolated unit.

v) Categories and the Comparison of Industrial Structures - a basic question in the study of industrial evolution and change is to ask if the processes at work modifying one industry can be used in any way to infer information about the forces influencing another; if we understand the dynamics that shape, say, the chemical industry, can that knowledge be used to understand, for instance, the evolution of the automotive industry? In order to answer such a question, we must have a systematic procedure for comparing the industries and a means for deciding whether they are abstractly equivalent. The (M,R)-system framework provides a means for making such comparisons through the mathematical apparatus termed "category theory" (Eilenberg and MacLane 1945, MacLane 1972). Briefly, any collection of sets $A, B, C, \ldots$, such that to each ordered pair (A, B) we have another set $H(A, B)$, the mappings from A to B, is called a category, provided certain primitive assumptions are satisfied for the set of mappings $H(A, B)$. We will defer any technical discussion of these matters to another paper, but it is important here to observe that every (M,R)-system is a category, in which the objects Ω, Γ are the sets and the metabolic maps $H(\Omega, \Gamma)$ are the mappings of the category. Thus, every firm can be regarded as a category, and by extending the sets and the mappings, so can every industry. If we change the sets Ω, Γ and/or the mappings $H(\Omega, \Gamma)$ obtaining a different firm, then we have a new category, and the machinery of category theory allows us to compare the structures of the two categories by means of mappings called functors. Roughly speaking, a functor is a sort of dictionary allowing us to translate the structure of one category into that of another, and conversely. This is exactly the type of tool which is needed to compare one firm or one industry with another. The systematic exploitation of this idea in the context of industrial structural change within the above (M,R)-framework offers the promise of unlocking many key features responsible for the dynamics underlying the evolution and development of modern global industries.

ACKNOWLEDGEMENTS

It is a pleasure to acknowledge the benefit of numerous conversations with R. Rosen on the subject of (M,R)-systems. Most of the ideas presented here have their origins in his earlier papers, written from a biological vantage point. I have only added a few embellishments here and there, and translated the results into an industrial setting.

REFERENCES

Businaro, U., "Comparing Natural Evolution and Technical Innovation"' Futures, April 1982.
Businaro, U., "Applying the Biological Evolution Metaphor to Technological Innovation", Futures, December 1983.
Casti, J., Linear Dynamical Systems, Academic Press, New York, 1987.
Casti, J., Nonlinear System Theory, Academic Press, New York, 1985.
Eilenberg, S., and MacLane, S., General Theory of Natural Equivalence", Tran. Amer. Math.Soc., 58 1945, pp.231-294.
Eilenberg, S., Automata, Languages and Machines, vol A, Academic Press, New York, 1974.
Kalman, R., Falb, P. and Arbib, M., Topics in Mathematical System Theory, McGrawHill, New York, 1969.
MacLane, S., Categories for the Working Mathematican, Springer, New York, 1972.
Nelson, R. and Winter, S., An Evolutionary Theory of Economic Change, Harvard University Press, Cambridge, 1982.
Rosen, R., "A Relational Theory of Biological Systems, I, II", Bull. Math-Biophysics 20, 1958a, pp. 245-260, 1959, pp. 109-128.

PART C: LOCATION AND SPATIAL DYNAMICS

Chapter 14

Continuous Models of Spatial Dynamics

M.J. BECKMANN

1. The dynamics of physical phenomena is based on the principle that interaction takes place only in small neighbourhoods. In spatial economics, a much weaker principle applies: interaction decreases with distance. In a system of interdependent cities - as in a Central Place System - buying and selling, spending and investing occur in one place although the agents responsible may be located in some other place at a considerable distance.

Let c_{ij} be the propensity of a resident at location j to spend in location i. All distances' effects are assumed to be represented in terms of these c_{ij}: a large distance between i and j implies a low value of c_{ij}.

Let income at i be Y_i, and assume permanent autonomous expenditure by outside agents, e.g. government, to be A_i. Then the incomes generated by these autonomous expenditures satisfy the condition

$$Y_i = A_i + \sum_j c_{ij} Y_j \tag{1}$$

or using a vector and matrix notation

$$y = a + Cy$$

$$y = (I - C)^{-1} a . \tag{2}$$

An equilibrium solution exists provided that eigenvalues of the non-negative matrix C are absolutely less than unity. A sufficient condition for this is that the propensity of any location to spend in all other locations is less than 1:

$$\sum_{i} c_{ij} < 1. \tag{3}$$

The dynamics of the system may be assumed to be determined by the simple linear relationships that apply to small deviations from equilibrium

$$Y_i\,(t) = A_i\,(t) + \sum c_{ij}\,Y_j\,\,(t\text{-}1)$$

or

$$y\,(t) = a(t) + C\,y\,(t\text{-}1) \tag{4}$$

which by successive substitution yields

$$y = \sum_{n=0}^{t} C^n\,a\,(t\text{-}n) \tag{5}$$

where we have set

$$C^0 = I.$$

This (essentially) non-spatial model is stable in view of the condition on the aggregate propensity to spend (3).

To capture the spatial side of the dynamic processes, one must either introduce distance explicity, or reintroduce the principle that interaction occurs only at close distances. We give examples of both.

2. A dynamic trade model between spatially separated markets may be set up as follows:

Let q_i (p_i) be an excess demand function describing local consumption minus production in market i. We assume that

$$q_i' < 0.$$

Let x_{ij} denote shipments from i to j, and let k_{ij} be the transportation cost between i and j. At first we assume that shipments are fixed through long term contracts. These imply that net imports are given as

$$-e_i = \sum_j x_{ij} - x_{ij}. \tag{6}$$

In equilibrium net imports equal local excess demand. The margin of excess demand over net imports is then

$$q_i (p_i) - \sum_j (x_{ij} - x_{ij}).$$

Walrasian price adjustment means that for small deviations from equilibrium price increases are proportional to excess demand:

$$\dot{p}_i = \beta \cdot [q_i (p_i) + e_i] \tag{7}$$

where

$$\beta > 0$$

measures the speed of adjustment. Assume that the excess demand function has the special form

$$q_i (p) = a_i - bp_i. \tag{8}$$

Then

$$\dot{p}_i = \beta [a_i - bp_i + e_i] \tag{9}$$

Thus each market has its own price adjustment process described by

$$\dot{p}_i = c_i - \beta\ b\ p_i, \tag{10}$$

where

$$c_i = \beta\ (a_i + e_i).$$

Notice that in this model, all markets have the same effective speed of adjustment β.

Alternatively, suppose that prices are adjusted according to the deviation of local inventories from desired levels $\bar{s}_i$. Then

$$\dot{p}_i = -\mu\ (s_i - \bar{s}_i)$$

$$\ddot{p}_i = -\mu\ \dot{s}_i = \mu\ [\ q_i\ (p_i) + e_i]$$

$$\ddot{p}_i = \mu\ [\ a_i - bp_i + e_i\ ]\ . \tag{11}$$

The equilibrium solution is given by

$$0 = a_i - b\bar{p}_i + e_i$$

or

$$\bar{p}_i = \frac{a_i + e_i}{b}\ . \tag{12}$$

Define

$$P_i = p_i - \bar{p}_i$$

to be the deviation from the equilibrium price. In terms of P_i one has the homogeneous equation

$$\ddot{P}_i = -\mu\ b\ P_i \tag{13}$$

which is solved by

$$P_i = \cos \sqrt{\mu\, b} \cdot t \tag{14}$$

and $P_i = \sin \sqrt{\mu b} . t$. (15)

The general solution is

$$P_i\,(t) = A \cos (\omega\, t + \alpha) \tag{16}$$

with

$$\omega = \sqrt{\mu\, b}\,.$$

This spatial market system exhibits undamped oscillations in the form of sine waves.

Depending on the phase differences between locations, these oscillations, although in principle unrelated, may still show a pattern of apparent "price waves" as proposed by August Lösch (1954, p. 267).

To study the possibility of non-accidental price waves one must resort to a description of spatial markets in continuous terms.

3. A continuous spatial market is described by prices as a function of two spatial coordinates x_1, x_2,

$$P\,(x_1,\, x_2)$$

and by a vector field $\phi(x_1,\, x_2)$ of commodity flow, whose direction is the direction of the flow and whose strength is the quantity of the commodity that passes through an orthogonal unit cross section per unit of time.

In equilibrium, excess supply (i.e. negative excess demand) equals the yield or "divergence" of this flow field

$$-q(x_1, x_2) = \text{div}\,\Phi(x_1, x_2) = \partial\Phi_1/\partial x_1 + \partial\Phi_2/\partial x_2. \tag{17}$$

This constitutes a commodity balance condition.

Secondly, in equilibrium, profits from trade must be zero for trades that actually take place, and negative for "inefficient" trades i.e. those in unprofitable directions. This may be written as

$$\text{grad } P = k \frac{\Phi}{|\Phi|} \tag{18}$$

where $k = k(x_1, x_2)$ represents the cost of transporting a unit of the commodity over a unit distance, regardless of the direction. (Beckmann and Puu 1985).

Let p be the derivation from equilibrium price and assume that the adjustment of flow is proportional to its profitability. So,

$$\frac{\partial\Phi}{\partial t} = a \text{ grad } p. \tag{19}$$

The adjustment of price is proportional to excess demand:

$$\frac{\partial p}{\partial t} = b[q + \text{div}\,\Phi] \,. \tag{20}$$

Differentiating (20) with respect to time

$$\frac{\partial^2 p}{\partial t^2} = b \frac{\partial q}{\partial P} \cdot \frac{\partial p}{\partial t} + b \text{ div } \frac{\partial \Phi}{\partial t}.$$

Now substitute for $\partial \Phi / \partial t$ from (19) and use the dot notation for derivatives with respect to time, so that

$$\ddot{p} = bq'\dot{p} + ab\,\Delta p \tag{21}$$

where

$$\Delta = \text{div grad} = \partial^2/\partial x_1^2 + \partial^2/\partial x_2^2$$

In particular, if excess demand is independent of price (as in the simple continuous transportation model)

$$\ddot{p} = ab\,\Delta p \tag{22}$$

and this is the wave equation in its simplest form.

We will not discuss the resulting price waves since their space and time profile depends crucially on the shape of the region. We will merely demonstrate the persistence of fluctuations. Multiplying (21) by $2\dot{p}$ one has

$$\partial p^2/\partial t = 2p\dot{p} = 2\,ab\,\dot{p}\Delta p. \tag{23}$$

Now

$$\dot{p}\ \text{div grad}\ p = \text{div}\,(\dot{p}\ \text{grad}\ p) - \text{grad}\ \dot{p}\ \text{grad}\ p. \tag{24}$$

Integrating (24) over the region A and applying the Gauss integral theorem

$$\int\int_A \dot{p}\ \text{div grad}\ p\ dx_1dx_2 = \int\int_A \text{div}\,(\dot{p}\ \text{grad}\ p) - \text{grad}\ \dot{p}\ \text{grad}\ p\ dx_1dx_2$$
$$= \int \dot{p}(\text{grad}\ p)_n\ dt - \int\int_A \text{grad}\ \dot{p}\ \text{grad}\ p\ dx_1\,dx_2\,.$$

Assume that the boundary conditions grad $p_n = 0$ on ∂A remain satisfied so that cross flows do not change. Then, using (19)

$$0 = a(\text{grad}\ p)_n = \partial\phi_n/\partial t = (\partial\phi/\partial t)_n\ .$$

Therefore, the line integral vanishes and we have

$$\int\int \dot{p} \text{ div grad } p \, dx_1 \, dx_2 = -\int\int \text{grad } \dot{p} \text{ grad } p \, dx_1 \, dx_2 \tag{25}$$

$$= -\frac{1}{2}\frac{d}{dt}\int\int (\text{grad } p)^2 \, dx_1 \, dx_2.$$

Integrating (23) over the region A and substituting (25)

$$d/dt \int\int \dot{p}^2 + ab \, |\text{grad } p|^2 \, dx_1 \, dx_2 = 0$$

or

$$\int\int \dot{p}^2 + ab \, |\text{grad } p|^2 \, dx_1 \, dx_2 = \text{constant.} \tag{26}$$

If the system was not initially in equilibrium, then the constant is positive. If the system ever passes through equilibrium so that $|\text{grad } p|^2 = 0$ then

$$\int\int \dot{p}^2 \, dx_1 \, dx_2 = \text{constant} > 0$$

so that the motion continues. The system can never settle down to equilibrium. It is unstable.

In the case where excess demand

$$q = q(p)$$

is a strictly decreasing function of price, the same argument leads to

$$d/dt \int\int \dot{p} + ab \, |\text{grad } p|^2 \, dx_1 \, dx_2 = \int\int aq' \, p^2 \, dx_1 dx_2 < 0 \tag{27}$$

since $q' < 0$ everywhere.

Thus price dependent demand acts as a damping factor in regard to the fluctuations of p in time and space. Prices converge to equilibrium.

Various modifications of this simple model may be explored. Essentially, price and flow adjustment through interaction with the immediate neighbourhood always leads to partial differential equations of the wave equation type, damped or undamped.

In the next section we return to the expenditure type of interaction now restricted to expenditures within the immediate neighbourhood.

4. Let $y(x_1,x_2,t)$ represent wealth in terms of liquid assets in locations x_1,x_2 at time t. Assume that spending units tend to spend this money at their own and in adjacent locations $x_i \pm \Delta x_i$ at time $t + \Delta t$. In discrete terms

$$y(x_1,x_2,t + \Delta t) = c_0 y(x_1,x_2,t) + c_1[y(x_1 - \Delta x_1,x_2,t) + y(x_1 + \Delta x_1,x_2,t) \tag{28}$$
$$+ y(x_1,x_2 - \Delta x,t) + y(x_1,x_2 + \Delta x,t)]$$

where c_0 is the propensity to spend at home and c_1 is the propensity to spend in any of the adjacent locations. Rewriting the bracket

$$[y(x_1 + \Delta x,x_2) - y(x_1,x_2)] - [y(x_1,x_2) - y[x_1 - \Delta x,x_2]$$
$$+[y(x_1,x_2 + \Delta x) - y(x_1,x_2)] - [y(x_1,x_2) - y(x_1,x_2 - \Delta x)] + 4y(x_1,x_2)$$

one observes

$$[y(x_1,x_2,t + \Delta t) - y(x_1,x_2,t)] / \Delta t$$

$$= [(c_0 + 4c_1 - 1)y(x_1,x_2,t)] / \Delta_t$$

$$+ \frac{\Delta x^2}{\Delta t} \frac{c_1}{\Delta x} [\frac{\Delta_1 y(x_1 + \Delta x,x_2)}{\Delta x} - \frac{\Delta_1 y(x_1,x_2)}{\Delta x}]$$

$$+ \frac{\Delta x^2}{\Delta t} \cdot \frac{c_1}{\Delta x} [\frac{\Delta_2 y(x_1,x_2 + \Delta x)}{\Delta x} - \frac{\Delta_2 y(x_1,x_2)}{\Delta x}] .$$

Now assume that the total propensity to spend equals unity:

$$c_0 + 4c_1 = 1. \qquad (29)$$

Then the first term cancels. Let the ratio of the time change to the space change be fixed at

$$\Delta x^2 / \Delta t = m/c_1 = \text{constant}. \qquad (30)$$

Going to the limit we obtain

$$\partial y/\partial t = m \Delta y$$

or

$$\partial y/\partial t = m[(\partial^2 y/\partial x^2{}_1) + (\partial^2 y/\partial x^2{}_2) \qquad (31)$$

the well-known diffusion equation.

This equation may be given the following interpretation. Introduce the gradient field grad y. We observe that money flow in the region is then described by

$$\Phi = -\text{grad } y \qquad (32)$$

and the change in liquid wealth is the divergence of this flow field. Thus liquid wealth can change only through the net yield of money flows

$$\dot{y} = -\text{div } \Phi = -m \text{ div}(-\text{grad } y) = m \Delta y.$$

Add now boundary condition $\Phi_n = 0$ on ∂A. Then it follows at once from the Gauss integral theorem

$$0 = \int \Phi_n \, ds = \int\int \text{div } \Phi \, dx_1 \, dx_2 = -(1/m \; d/dt) \int\int y \, dx_1 \, dx_2 .$$

Total liquid wealth or money stock must remain unchanged.

Consider now an equal distribution of liquid wealth throughout the region

$$y(x_1 x_2, t) = \bar{y}. \tag{33}$$

We show that this is the only stable equilibrium. Consider

$$1/2 \ d/dt \int\int (y-\bar{y})^2 dx_1 \ dx_2 = \int\int (y-\bar{y}) \cdot \dot{y} \ dx_1 dx_2$$

$$= m\int\int (y-\bar{y})\Delta y \ dx_1 dx = m\int\int (y-\bar{y}) \text{div grad } y \ dx_1 dx_2.$$

Now

$$(y-\bar{y}) \text{ div grad } y = \text{div}[(y-\bar{y}) \text{ grad } y] - \text{grad } y \text{ grad } y.$$

Also by the Gauss integral theorem

$$\int\int \text{div}[(y-\bar{y})\text{grad } y] \ dx_1 dx_2 = -\int (y-\bar{y})\Phi_n \ ds = 0$$

by the boundary condition, using (32).

Substituting

$$d/dt \int\int (y-\bar{y})^2 \ dx_1 dx_2 = -m\int\int (\text{grad } y)^2 \ dx_1 dx_2 < 0 \tag{34}$$

proves the assertion.

Suppose now that spending by residents of a location is not financed entirely by net receipts from trade, but by outside earnings as well. Let $z(x_1, x_2)$ be the density of these net earnings which may be negative. In that case, residents of this location pay wages to outsiders, presumably out of earnings from trade. Let the expenditure or money flow equation be as follows:

$$\partial y/\partial t = z - m \text{ div}(-\text{grad } y). \tag{35}$$

Constancy of aggregate disposible wealth or aggregate liquid assets implies

$$0 = \int\int (\partial y/\partial t) \ dx_1 dx_2 = \int\int z \ dx_1 \ dx_2 + m\int (\text{grad } y)_n \ ds = \int\int z \ dx_1 \ dx_2. \tag{36}$$

Thus aggregate net earnings z from outside locations must add up to zero.

Now let y^* be the stationary solution of that system

$$z + m \text{ div grad } y^* = 0. \tag{37}$$

Consider

$$(1/2\ d/dt) \int\int (y - y^*)^2\, dx_1\, dx_2 =$$

$$\int\int (y - y^*)\,(dy/dt)\; dx_1\, dx_2 = \int\int (y - y^*)(z + m \text{ div grad } y) dx_1\, dx_2$$

$$= m\int\int (y - y^*) \text{ div grad}(y - y^*) dx_1\, dx_2$$

using (35) and (37)

$$= m\int\int \text{div}[(y - y^*)\text{grad}(y-y^*)]\; dx_1\, dx_2 - m\int\int |\text{grad}(y-y^*)|^2\, dx_1\, dx_2 < 0 \tag{38}$$

since the first integral vanishes by the Gauss integral theorem and the boundary condition. This shows the stability of the equilibrium solution y^* of (37).

This paper has presented some examples of economic processes in space, whose dynamics are described by the well known wave and diffusion equations. These do not by any means exhaust the existing possibilities. In particular, we have not attempted to investigate the emergence of singularities or "catastrophies" which are often associated with spatial growth processes, as in the phenomenon of growth poles. These interesting topics are beyond the scope of the present paper.

REFERENCES

Beckmann, M. and Puu, T, Spatial Economics: Density, Potential and Flow, North Holland, Amsterdam, 1985.

Lösch, A., The Economics of Location, Yale University Press, New Haven, 1954.

Chapter 15

The Onset of Turbulence in Discrete Relative Multiple Spatial Dynamics

D. S. DENDRINOS AND M. SONIS

THE RELATIVE FIXED POINT PROCESS

The iterative algorithm (mapping) of relative discrete dynamics to be discussed here was first presented in a paper by Sonis and Dendrinos (1984), where the two- location, one-stock problem was addressed. Its connections with fundamental socio-spatial dynamics are expanded in Dendrinos and Sonis (1987), where the full gamut of I-location, J-stock models are presented together with some empirical evidence.

Multiple-location, one-stock discrete dynamics of the form:

$$x_i(t+1) = (A_i F_i) / (\sum_j A_j F_j) \quad ; \quad i=1,2,...,I \tag{1}$$

$$A_i > 0 \quad ; \quad i=1,2,...,I \tag{2}$$

$$F_i = F_i[x_i(t) \quad ; \quad i=1,2,...,I] > 0 \quad ; \quad i=1,2,...,I \tag{3}$$

$$0 < x_i(t) \quad , \quad x_i(t+1) < 1 \quad : \quad i=1,2,...,I \quad ; \quad 0 \leqq t \leqq T \; , \tag{4}$$

are discussed in this paper; in particular, we present results obtained when I=2, and I=3. For higher dimensional problems, and for cases with more than one iteration (time period) lags, e.g.

$$x_i(t+n) = A_i F_i / \sum_j A_j F_j \quad ; \quad n > 1 \tag{5}$$

see Dendrinos and Sonis (1987). In particular, the following log-linear specifications of the F-functions are analysed:

$$F_i = \prod_{k=1}^{I} x_k(t)^{\alpha_{ik}} \tag{6}$$

$$-\infty \leqq \alpha_{ik} \leqq +\infty \quad ; \quad i,k=1,2,...,I. \tag{7}$$

Exponential formulations of the F-functions:

$$F_i = \exp \prod_{k=1}^{I} x_k(t)^{\alpha_{ik}} \tag{8}$$

are also discussed in Dendrinos and Sonis (1987) where it is demonstrated that regardless of the values of the bifurcation parameters, A_i, and the exponents, α_{ik}, the results are always fixed points in the case of the exponential formulation; for any dimension $I > 0$. These results are obtained analytically.

Since the pioneering work by Lorenz (1963), and the subsequent landmark findings by May (1976) and Feigenbaum (1978) on simple discrete iterative processes, a great deal of interest has been attracted to the field of bifurcation theory (see, for instance, Haken, 1983). Developments in this field have been either analytical or computational, relying increasingly on numerical simulations as the forms examined become more complex. The numerical simulations are creating a new subdiscipline within the field, known as experimental mathematics; see S. Ulam (1985).

Two of the central issues in this emerging field are the questions of "faithfulness" and "completeness" . "Faithfulness" questions whether the results obtained by the numerical simulations are due to the intrinsic processes of the map, the peculiarities of the algorithms employed, or the computer's accuracy. "Completeness" questions whether the numerical experiments, no matter how extensive, can always provide the full richness of the mappings under examination.

This paper presents a few unexpected, interesting results, previously unidentified in the mathematical literature (to our knowledge). They contain significant implications for the evolution of general (abstract) socio-spatial systems, which are elaborated in Dendrinos and Sonis (1987). However, a few examples of the use of the stock-location universal algorithm are also outlined here.

One area where this discrete map may be of direct application is that of regional population dynamics expressed in relative terms. Interdependencies among population stocks, over space and time, can be driven by comparative advantages enjoyed by the various locations competing for population(depicted by the model's parameters). Fast adjustments in population stocks are dictated by the current values of these parameters and their (slow) changes.

Another area of possible application is that where capital stocks and population (labor) stocks interact: inter-stock, inter-spatial, inter-temporal, interdependencies and competition can be incorporated through the use of the universal map. Different speeds of change in these stocks can be accounted for by introducing different time lags.

Finally, one could extend this analysis by treating distance similarly to time (i.e., as an iterative process where temporal and spatial ticking occurs) where both are (simultaneous) determinants of stock dynamics. Clearly, research along these lines holds considerable future promise.

THE FUNDAMENTAL BIFURCATION IN DISCRETE DYNAMICS: THE TWO-LOCATION, ONE-STOCK CASE

In proceeding to demonstrate the fundamental bifurcation in discrete (iterative) dynamics – the equivalent of Hopf's bifurcation in continuous dynamics – we make use of the following specifications from (6,7):

$$x_1(t+1) = F_1 / (F_1 + AF_2) = 1/(1 + AF) \quad (9)$$

$$x_2(t+1) = 1 - x_1(t+1) = AF/(1 + AF) \quad (10)$$

$$F = x_1(t)^{\alpha} x_2(t)^{\beta} > 0 \quad (11)$$

$$0 < x_1(t) \ , \ x_2(t) < 1 \quad (12)$$

$$A > 0 \ , \quad (13)$$

so that at any time period t:

$$[x_2(t+1)] / [x_1(t+1)] = [1 - x_1(t+1)] / [x_1(t+1)] = [x_2(t+1)] / [1 - x_2(t+1)] = AF. \quad (14)$$

Formulating the corresponding slopes, one obtains:

$$\begin{aligned} s_{11}(t+1) &= \partial x_1(t+1)/\partial x_1(t) = -x_1(t+1)^2 A[\partial F/\partial x_1(t)] \\ &= -x_1(t+1)^2 A[\alpha F/x_1(t) - \beta F/x_2(t)] \quad (15) \\ &= -[\alpha x_2(t) - \beta x_1(t)] \ x_1(t+1)x_2(t+1)/x_1(t)x_2(t) \end{aligned}$$

which at equilibrium (x_1^*, x_2^*) yields

$$s_{11}^* = [\alpha x_2^* - \beta x_1^*] \ . \quad (16)$$

Similarly:

$$\begin{aligned} s_{12}(t+1) &= \partial x_1(t+1)/\partial x_2(t) = -x_1(t+1)^2 A[\partial F/\partial x_2(t)] \\ &= -x_1(t+1)^2 A[\beta F/x_2(t) - \alpha F/x_1(t)] \\ &= -[\beta x_1(t) - \alpha x_2(t)] \ x_1(t+1)x_2(t+1)/x_1(t)x_2(t) \\ &= -s_{11}(t+1) \quad (17) \end{aligned}$$

$$s_{12}^* = \alpha x_2^* - \beta x_1^* \quad (18)$$

$$s_{21}(t+1) = \partial x_2(t+1)/\partial x_1(t) = x_1(t+1)x_2(t+1)[\partial F/\partial x_1(t)]$$

$$= [\alpha x_2(t) - \beta x_1(t)]\ x_1(t+1)x_2(t+1)/x_1(t)x_2(t) = - s_{11}(t+1) = s_{12}(t+1) \tag{19}$$

$$s^*_{21} = \alpha x^*_2 - \beta x^*_1 \tag{20}$$

$$s_{22}(t+1) = \partial x_2(t+1)/\partial x_2(t) = x_1(t+1)x_2(t+1)[\partial F/\partial x_2(t)]$$

$$= -[\alpha x_2(t) - \beta x_1(t)]\ x_1(t+1)x_2(t+1)/x_1(t)x_{2(t)} = -s_{21}(t+1) = s_{11}(t+1) \tag{21}$$

$$s^*_{22} = -[\alpha x^*_2 - \beta x^*_1]\ . \tag{22}$$

In the matrix of slopes at equilibrium, the Jacobi matrix J* is given by:

$$J^* = \begin{bmatrix} s^*_{11} & s^*_{12} \\ s^*_{21} & s^*_{22} \end{bmatrix} = \begin{bmatrix} -\alpha x^*_2 & -\beta x^*_1 \\ \alpha x^*_2 & \beta x^*_1 \end{bmatrix} \tag{23}$$

We note that: det J* = 0, or:

$$s^*_{11}\ s^*_{22} = s^*_{12}\ \ s^*_{21} = -\alpha\beta x^*_1\ x^*_2 \tag{23.1}$$

and that its trace TrJ* is:

$$s^*_{11} + s^*_{22} = \beta x^*_1 - \alpha x^*_2. \tag{23.2}$$

Forming the characteristic polynomial of the matrix J* one obtains the real eigenvalue λ from:

$$\lambda = \text{TrJ}^* = s^*_{11} + s^*_{22} = (\alpha + \beta)\ x^*_1 - \alpha. \tag{24}$$

According to the well-known von Neumann theorem for difference equations, (see Saaty 1967, p. 168) the equilibrium state (x^*_1, x^*_2) will be stable if $|\lambda| < 1$. When the eigenvalue $|\lambda| = 1$, then bifurcations of three different types occur. They shall be discussed next making use of numerical simulations.

Two thresholds determine the nature of these bifurcations: $\eta_{+1} = (\alpha + 1)/(\alpha + \beta)$ associated with the value $\lambda = 1$, and the threshold $\eta_{-1} = (\alpha - 1)/(\alpha + \beta)$ associated with the value $\lambda = -1$. In the parameter space where $\alpha + \beta < 0$, the three bifurcations manifest themselves as follows: if $0 \leqq \eta_{-1} \leqq 1$ and $\eta_{+1} < 0$ then as A moves from zero to $+\infty$ only one type of bifurcation occurs in which x* switches from a fixed point into a stable two-period cycle. It takes place in the area $-1 \geqq \beta$, $-1 \leqq \alpha \leqq 1$. The second type of bifurcation occurs, transforming x* from an unstable equilibrium to an attractor, as A moves from zero to $+\infty$ when $0 \leqq \eta_{+1} \leqq 1$ and $\eta_{-1} > 1$; it occurs in the area $-1 \leqq \beta \leqq 1$, $\alpha \leqq -1$. Finally, the third type of bifurcation occurs when $0 \leqq \eta_{\pm 1} \leqq 1$; now as A moves from zero to $+\infty$ the equilibrium is transformed from unstable, to stable, to period doubling cycles, and to chaos. This phenomenon occurs in the parameter space $\alpha < 1$, $\beta < 1$.

Bifurcation type one is the discrete dynamics equivalent of the Hopf bifurcation in continuous maps. Bifurcation type two transforms a competitive exclusion (unstable equilibrium) through a one-sided stable fixed point, to two equilibria, one stable and the other unstable (competitive exclusion) - the final result depending on the starting value for x. Bifurcation type three is the one widely studied in reference to the discrete logistic model.

When $\alpha + \beta > 0$ then the first type of bifurcation transforming a fixed point to a stable two-period cycle, as A moves from $+\infty$ to zero, occurs when $0 \leqq \eta_{-1} \leqq 1$ and $\eta_{+1} > 1$. It takes place in the mirror image (in reference to the $\alpha + \beta = 0$ axis), part of the parameter space where this bifurcation occurs under the previous case ($\alpha + \beta < 0$); specifically, it is found where $\alpha \geqq 1$ and $-1 \leqq \beta \leqq 1$. The second type of bifurcation transforming an unstable equilibrium into a fixed point (attractor), occurs when $0 \leqq \eta_{+1} \leqq 1$ and $\eta_{-1} < 0$ corresponding to the parameter space $-1 \leqq \alpha \leqq 1$ and $\beta \geqq 1$. The third type of bifurcation occurs when $0 < \eta_{\pm 1} \leqq 1$ corresponding to the parameter space $\alpha \geqq 1$, $\beta \geqq 1$.

If the bifurcation points are found in the value $A = A^*$, in the $\alpha + \beta < 0$ space, then the corresponding bifurcations in the $\alpha + \beta > 0$ are found in $A = A^{**} = 1/A^*$. Specifically, the type one or type two bifurcations occur at points:

$$A^* = (1 - x_1^*)^{(1-b)} / x_1^{*(1+\alpha)} = 1/A^{**}. \tag{25}$$

A more detailed presentation of the above is found in Sonis and Dendrinos (1984) and Dendrinos and Sonis (1987).

In summary, there are three fundamental bifurcation types: (a) The fundamental bifurcation which is equivalent to the Hopf bifurcation in continuous dynamics, and transforms a stable attractor (fixed point), through a center, to a stable two-period cycle in discrete relative mappings; (b) the fundamental bifurcation transforming competitive exclusion to an attractor, through a one-sided stable equilibrium; (c) the fundamental bifurcation which is equivalent to the May bifurcation of the map $y(t+1) = ay(t)\,[1-y(t)]$, and in discrete relative mappings transforms an unstable equilibrium, to an attractor, to period doubling cycles, and then to chaos. The necessary condition for the presence of the May type bifurcation is the collapse of the two thresholds η_{+1} and η_{-1} into the space (0,+1); in other words, the simultaneous presence of type one and type two bifurcations.

For the value of A^* obtained from (25), at which the above holds precisely, the iterative process is at an orbital motion: given any starting value, there is a two-period cycle which involves the two starting values. At values different than A^*, the iterative process converges to a fixed point (x_1^*, x_2^*) so that for t odd:

$$\lim_{t\to\infty} x_1(t) = x_1^* \quad ; \quad \lim_{t\to\infty} x_1(t+1) = x_1^* \tag{26.1}$$

$$\lim_{t\to\infty} x_2(t) = x_2^* \quad ; \quad \lim_{t\to\infty} x_2(t+1) = x_2^* \tag{26.2}$$

or it converges to a stable two-period cycle:

$$\lim_{t\to\infty} x_1(t) = x_1^*(1) \quad ; \quad \lim_{t\to\infty} x_1(t+1) = x_1^*(2) \tag{27.1}$$

$$\lim_{t\to\infty} x_2(t) = x_2^*(1) \quad ; \quad \lim_{t\to\infty} x_2(t+1) = x_2^*(2) \tag{27.2}$$

so that

$$x_1(t) + x_2(t) = 1 \tag{28}$$

$$x_1(t+1) + x_2(t+1) = 1 \tag{29}$$

$$x_1^* + x_2^* = 1 \tag{30}$$

$$x_1^*(1) + x_2^*(1) = 1 \tag{31}$$

$$x_1^*(2) + x_2^*(2) = 1. \tag{32}$$

In the case where A = A* and $\alpha + \beta < 0$, so that $s_{11}^* = -1$, the fixed point undergoes a transition: at this critical value any starting point $\{x_1(0), x_2(0)$ so that $x_1(0) + x_2(0) = 1\}$ is a part of an unstable two-period cycle. This is equivalent to the center type dynamic equilibrium in the continuous case. Thus, a fixed point is transformed through an orbit to a stable limit cycle in the form of a stable two-period cycle; see Figure 1. The iterative behavior is shown in Figure 2. For a starting value outside the ranges of $x_1^*(1)$, $x_1^*(2)$ the process converges toward these limits from the outside; whereas, for starting values within these thresholds the process converges toward these limits from the inside.

A numerical example is given below for the case where $\alpha = 1.5$, $\beta = -.4$, as in Figure 2 and Table 1. The cases of odd and even iteractions are exposed, when 500 iterations are carried out. For values of A up to the bifurcation point A*, found somewhere in the range 3.05 < A* < 3.10, $x_1(t)$ declines and $x_1(t+1)$ increases – both converging toward x^*_1 as t approaches $+\infty$. Beyond this critical value $x_1(t)$ increases and $x_1(t+1)$ declines as $x_1(t)$ converges to $x_1^*(1)$, whereas $x_1(t+1)$ converges toward $x_1^*(2)$. This allows one to detect, given any number of iterations – and not neccessarily when $T >> 0$ – whether A lies before or after the critical value A*. Up till A* and for ranges close to this threshold, pseudo-cycles are formed, which are eliminated as $T \to +\infty$. Note also that up till A* the mean value equation $x_1(t) = [x_1(t) + x_1(t+1)] / 2$ is decreasing, whereas it increases afterward for odd iterates.

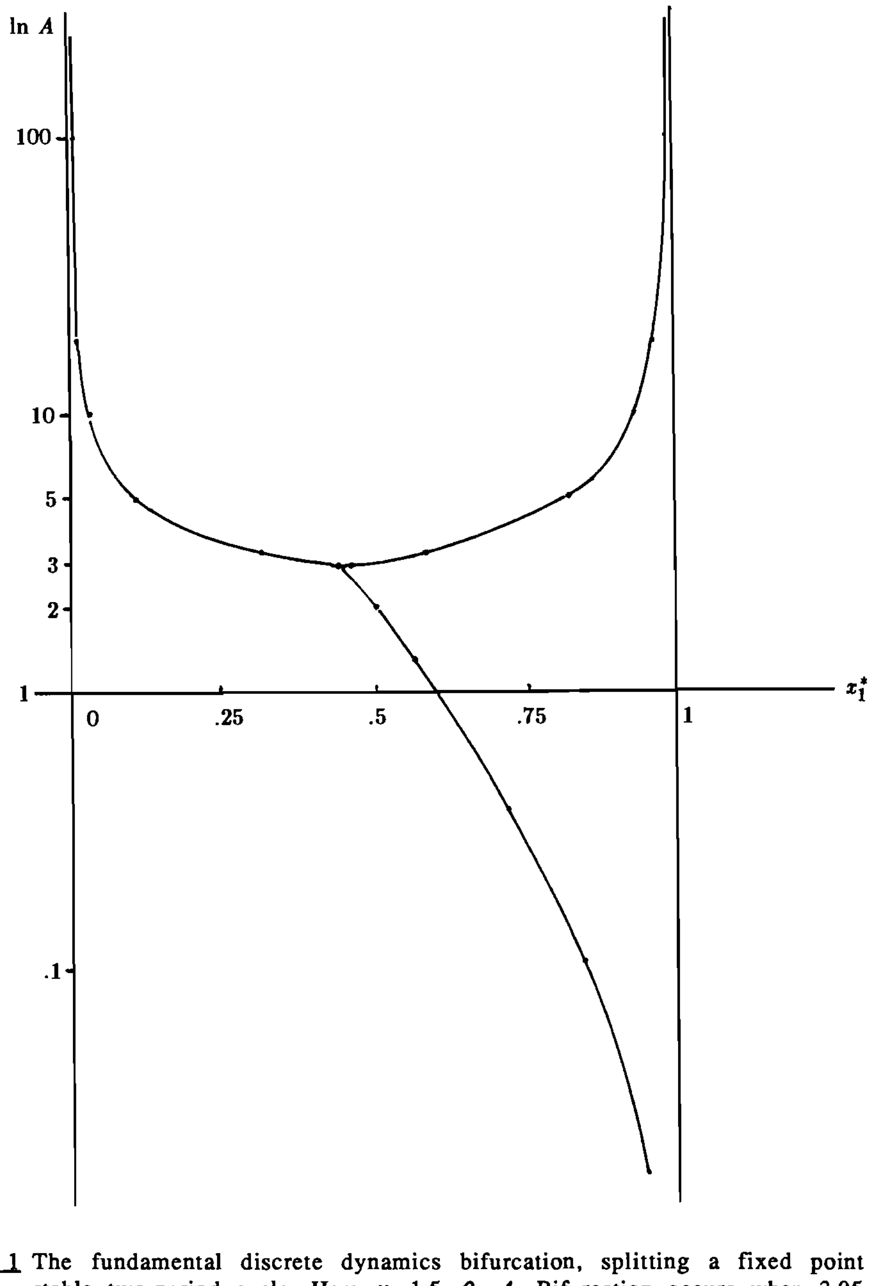


Figure 1 The fundamental discrete dynamics bifurcation, splitting a fixed point into a stable two-period cycle. Here α=1.5, β=-4. Bifurcation occurs when 3.05 < A < 3.10 where x* = .4545....

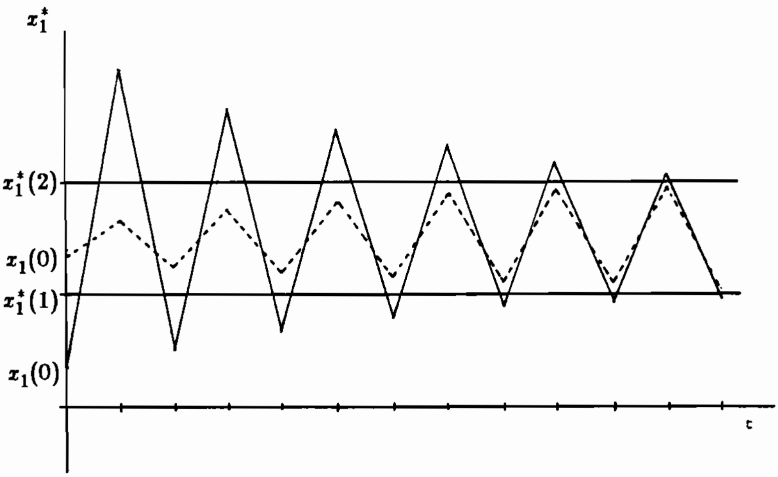


Figure 2 The stable two-period cycle: the discrete dynamics equivalent to the stable limit cycle of continuous dynamics. Points $x_1(0)$ identify two starting values, both cases converging to the limits $x_1^*(1)$ and $x_1^*(2)$ either from the inside (-----) or the outside (———).

Table 1 Numerical experiments for the log-linear specification of the two-locations, one-stock iterative dynamics. The parameters generating these results are: $\alpha = 1.5$, $\beta = -.4$; results are based on 500 iterations. A* identifies the critical value where the Hopf-like discrete dynamics bifurcation occurs.

A	$x_1^*(1)$	$x_1^*(2)$	$x_1^*(1)+x_1^*(2)$
1.00	.5993		
2.00	.5089		
2.25	.4938	same as	
2.50	.4804	in $x_1^*(1)$	
2.75	.4684		
2.90	.4610	.4624	.9234
2.95	.4573	.4620	.9193
3.00	.4501	.4650	.9151
3.05	.4337	.4777	.9114
A*	. . . . .	. . . . .	. . . . .
3.10	.4021	.5074	.9095
3.25	.3164	.5975	.9139
3.75	.2111	.7144	.9255
4.00	.1821	.7480	.9301

2. LOCAL AND PARTIAL TURBULENCE: THE THREE-LOCATION, ONE-STOCK CASE

This event, depicted in Figure 3, identifies a fixed point in one variable and a stable two-period cycle in the other two. It is referred to as "local" because only two variables out of the three in this case are in an oscillatory motion. It is called "partial" because the cycles are periodic. The full gamut of turbulence, as found in other circumstances (in this and other discrete dynamic processes), is not replicated during this event.

Formulating the three-location, one-stock problem of the universal discrete relative dynamics model one has:

$$x_i(t+1) = A_i F_i / \sum_j A_j F_j \quad , \quad i=1,2,3 \tag{33}$$

where F is specified in a log-linear manner:

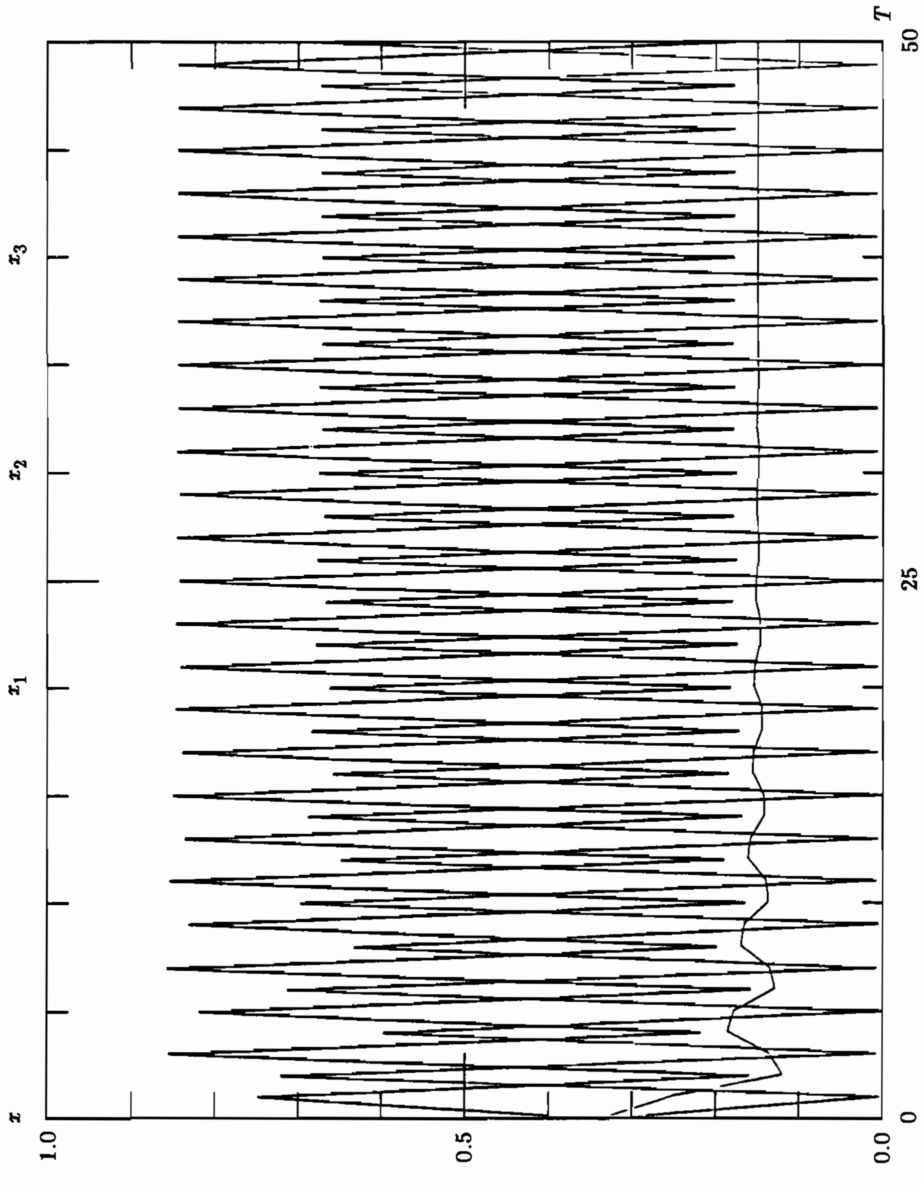
$$F_i = \prod_k x_k^{\alpha_{ki}} \quad , \quad i,k=1,2,3. \tag{34}$$

Without loss of generality one may assume: $A_1 = 1$, $\alpha_{1i} = 0$ for i=1,2,3. Then the above becomes:

$$\begin{aligned} x_1(t+1) &= 1 / (1+K) \\ x_2(t+1) &= F_2 / (1+K) \\ x_3(t+1) &= F_3 / (1+K) \end{aligned} \tag{35}$$

where $K = A_2 \prod_h x_h^{\alpha_{2h}} + A_3 \prod_h x_h^{\alpha_{3h}}$.

Formulating the Jacobian $s_{ij}(t+1) = \partial x_i(t+1) / \partial x_j(t)$ when i,j=1,2,3 one obtains:



<u>Figure 3</u> The local and partial turbulence phenomenon; the parameter values responsible for this simulation are:
$A_1=1$, $A_2=10^{-4}$, $A_3=1$; $\alpha_{11}=\alpha_{12}=\alpha_{13}=0$; $\alpha_{21}=\alpha_{22}=\alpha_{23}=-1.5$; $\alpha_{31}=\alpha_{32}=0$, $\alpha_{33}=-1.0$.

$$M = \begin{bmatrix} s_{11}(t+1) & s_{12}(t+1) & s_{13}(t+1) \\ s_{21}(t+1) & s_{22}(t+1) & s_{23}(t+1) \\ s_{31}(t+1) & s_{32}(t+1) & s_{33}(t+1) \end{bmatrix} \tag{36}$$

with entries:

$$s_{1j}(t+1) = -x(t+1)^2(\partial K/\partial x_j(t)) \quad , \quad j=1,2,3; \tag{37.1}$$

$$s_{hi}(t+1) = x_1(t+1)\ [(\partial F_h/\partial x_j(t)) - x_1(t+1)\ F_h(\partial K/\partial x_j(t))] \quad h=2,3 \quad , \quad j=1,2,3. \tag{37.2}$$

$$\sum_{j=1}^{3} s_{ij}(t+1) = 0 \quad , \quad i=1,2,3. \tag{37.3}$$

In the log-linear case, one obtains:

$$\partial F_h/\partial x_j(t) = \alpha_{hi}\ F_h / x_j(t) \tag{38}$$

$$\partial K/\partial x_j(t) = (A_2\ \alpha_{2j} + A_3\ \alpha_{3j})\ F_j / x_j(t), \tag{39}$$

so that the entries of the Jacobian become:

$$s_{ij}(t+1) = -x_1(t+1)^2(A_2\,\alpha_{2j} + A_3\,\alpha_{3j})\ F_j / x_j(t) \quad ; \quad j=1,2,3 \tag{40}$$

$$s_{hi}(t+1) = x_1(t+1)\ (F_h/x_j(t))\ [\alpha_{hj} - x_1(t+1)\ (A_2\,\alpha_{2j} + A_3\ \alpha_{3j})\ F_j] \tag{41}$$

$$h=2,3 \quad , \quad j=1,2,3.$$

Again, the properties of the dynamics depend upon the relative size of s^*_{ii}, i=1,2,3. At equilibrium:

$$s^*_{11} = -x^{*2}_1\ (A_2\,\alpha_{21} + A_3\,\alpha_{31}) \tag{42}$$

$$s^*_{22} = x^*_1\ (F^*_2/x^*_2)\ [\alpha_{22} - x^*_1(A_2\,\alpha_{22} = A_3\ \alpha 32)\ F^*_2] \tag{43}$$

$$s^*_{33} = x^*_1\ (F^*_3/x^*_3)\ [\alpha_{22} - x^*_1\ (A_2\ \alpha_{23} + A_3\ \alpha_{32})\ F^*_3]\ . \tag{44}$$

For any combination of A_2, A_3, given appropriate exponents α_{ij}, the following conditions may result in the entries of the Jacobian: $|s^*_{11}| < 1$, $s^*_{jj} = -1$, j=2,3. At this point a state of three fixed points is transformed into a state of partial and local turbulence, where one location's abundance, say x^*_1, is a fixed point (attractor), whereas the other two locations' abundance x^*_2 and x^*_3 are in a stable two-period cycle as in Figure 3.

The case of local turbulence, where one location exhibits fixed point behavior, the other two sharing the stock in a stable two-period cycle, <u>is another special case of the fundamental bifurcation in discrete dynamics</u>. Under special conditions it could be shown that: $x^*_2(1) = x^*_3(2)$ and $x^*_2(2) = x^*_3(1)$, the socio-spatial system exhibiting local role reversal.

Numerical experimentation in the four-location, one-stock case indicates that we could observe one or two locations at a fixed point, the rest exhibiting a stable two-period cycle. Further, these results are also replicated in the two-stock, two-location case, where one stock could be at a fixed point in both locations, whereas the other exhibits a stable two-period cycle (at times involving role reversal, implying that $x^*_1(1) = x^*_2(2)$, $x^*_1(2) = x^*_2(1)$, whereas the second stock shows fixed point (y^*_1, y^*_2) behavior). Finally, numerical experimentation in the two-stock, two—location case indicates that for fixed exponents, as one of the bifurcation parameters varies smoothly, switching may occur, whereby the stocks' behavior is reversed: the stock showing fixed point behavior reverses to a stable two-period cycle, whereas the opposite happens to the second stock.

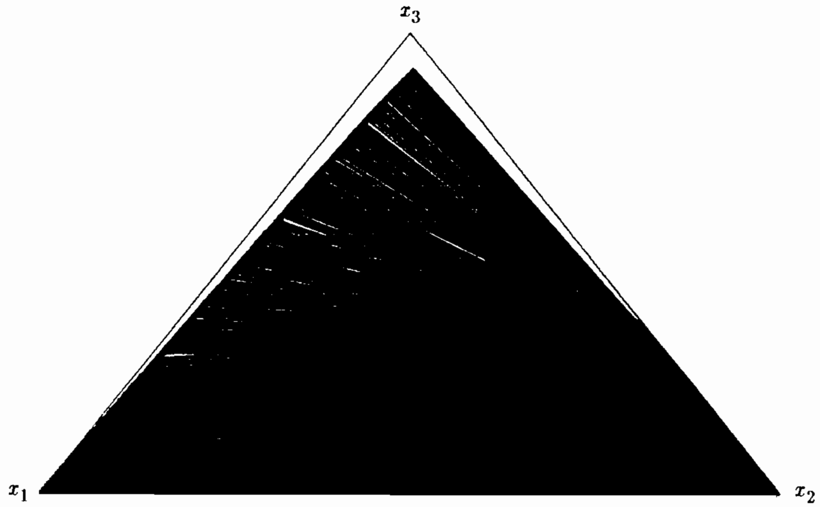
3. STRANGE CONTAINERS: THE THREE-LOCATION, ONE-STOCK CASE

For socio-spatial systems the chaotic behavior of dynamic processess may be of paramount interest, since that is where, in all likelihood, socio-spatial systems are found given the relative abundance of chaotic behavior in iterative discrete dynamics. In the three-dimensional case of the three-location, one-stock problem, the nature of the chaotic regime is not the same at all points in the parameter space where chaos prevails. Chaos, in fact, is not only deterministic in its intra-chaotic behavior at each point, but furthermore, it is changing in a deterministic manner during its inter-chaotic transformations.

There are regimes of chaos where the intra-chaotic motion spans almost the whole spectrum inside the triangle of behavior, giving rise to "global strange containers"; see Figure 4., Points in the parameter space also exist in this three-dimensional case where the chaotic motion is confined to "local strange containers" of various sizes. They range from large areas inside the triangle, shown in Figure 5, to relatively medium portions of the space, as in Figure 6. A "mini strange container" is shown in Figure 7. Although the numerical aspects of the transitions in the chaotic regimes have been partly unravelled, the analytical properties still remain intractable.

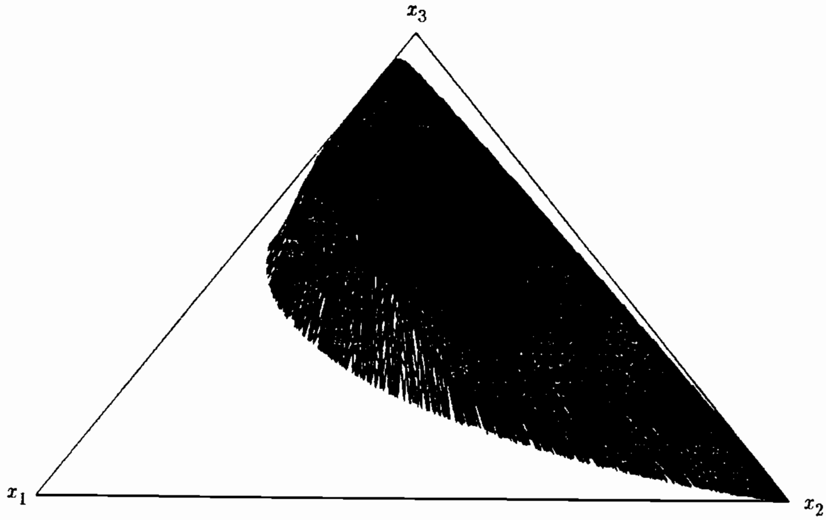
The term "container" has been used here to designate the opposite phenomenon of that implied by the notion of "attractor". Strange attractors are limits toward which dynamic paths converge; they outline areas in a envelope-type manner. Cases where "strange attractors" occur in the universal discrete dynamics model of three-location, one-stock case are shown in Figure 8. A "hybrid" case, where a combination of a strange container and a strange attractor occurs, is shown in Figure 9. All of the results shown are independent of the initial values of the mapping.

It is highly unlikely that these are the only unexpected and innovative events present in the universal discrete dynamics parameter space. Numerical experimentation may indeed reveal other phenomena present in this vast space, shedding much insight on the discrete spatial dynamics of social systems.



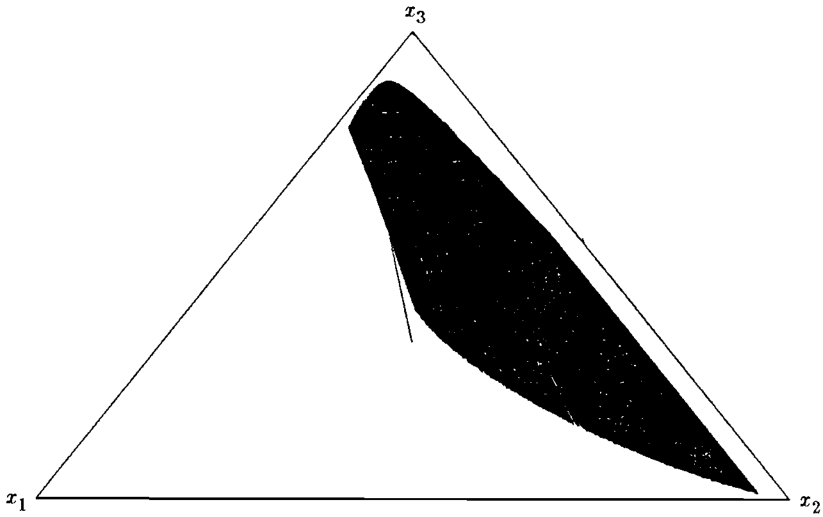
$$\begin{bmatrix} 0 & 0 & 0 \\ -1.5 & -1.5 & -1.5 \\ -1.5 & -1.0 & -0.5 \end{bmatrix}$$

Figure 4 A global "strange container". Parameter values responsible: $A_1=1$, $A_2=10^{-6}$, $A_3=10^{-4}$. 20,000 iterations.



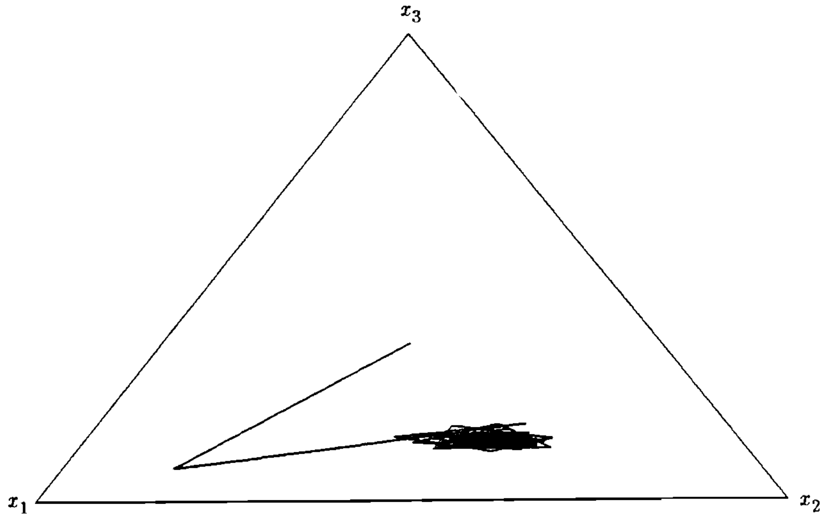
$$\begin{bmatrix} 0 & 0 & 0 \\ -1.5 & -1.5 & -1.5 \\ 0.5 & 0.5 & -0.5 \end{bmatrix}$$

Figure 5 A local "strange container" of large size. Parameter values responsible: $A_1=1$, $A_2=10^{-4}$, $A_3=10$. 20,000 iterations.



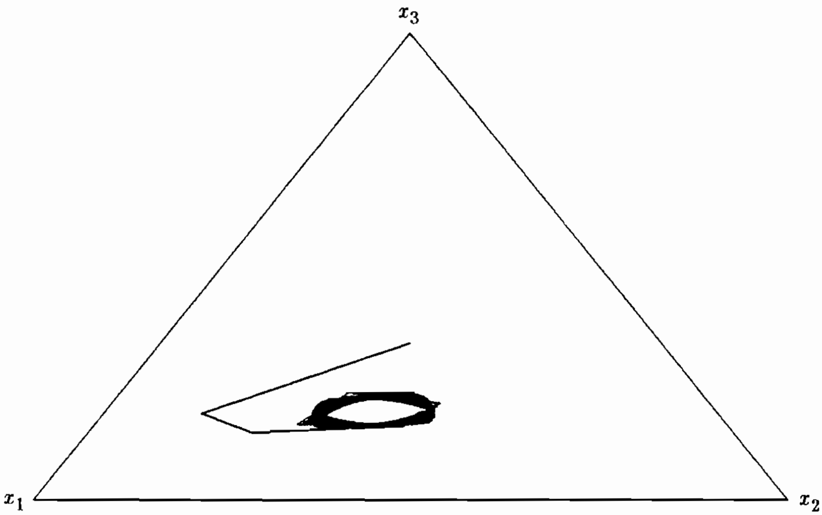
$$\begin{bmatrix} 0 & 0 & 0 \\ -1.5 & -1.5 & -1.0 \\ 0.5 & 0.5 & -0.5 \end{bmatrix}$$

Figure 6 A local "strange attractor" of medium size. Parameter values responsible: $A_1=1$, $A_2=10^{-3}$, $A_3=10$. 20,000 iterations. The line from the centre indicates the initial value.



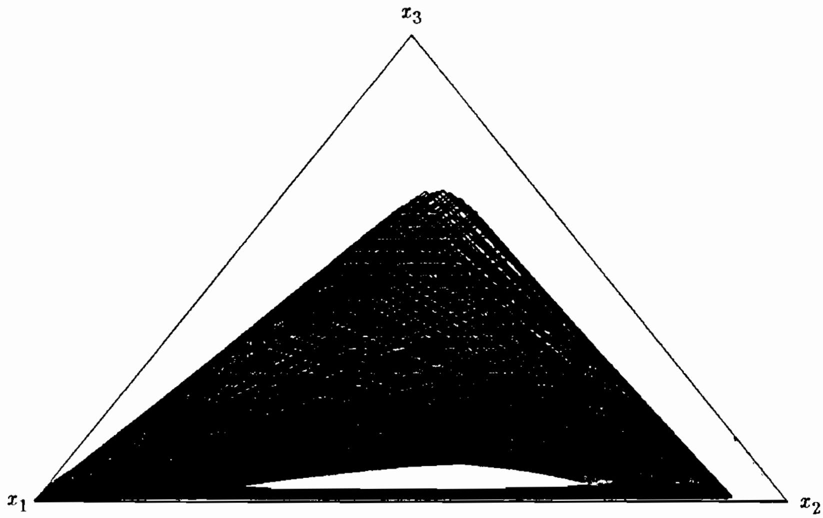
$$\begin{bmatrix} 0 & 0 & 0 \\ 1.5 & 1.5 & -1.5 \\ -0.5 & 0 & -1.5 \end{bmatrix}$$

Figure 7 A "mini strange container". Parameter values $A_1=A_2=1$, $A_3=10^{-2}$. 500 iterations



$$\begin{bmatrix} 0 & 0 & 0 \\ 1.5 & 1.5 & -1.5 \\ -1.5 & 0 & -1.5 \end{bmatrix}$$

Figure 8 Strange attractor. Parameter values $A_1=A_2=1$, $A_3=10^{-2}$. 10,000 iterations.



$$\begin{bmatrix} 0 & 0 & 0 \\ 1.5 & 1.5 & -1.5 \\ -1.5 & .5 & -1.0 \end{bmatrix}$$

Figure 9 A "hybrid" strange container and attractor. Parameter values responsible $A_1=1$, $A_2=.1$, $A_3=10^{-3}$. 20,000 iterations.

CONCLUSIONS

An overview of a few interesting new events in the simulation runs involving relative discrete two- and three-location, one-stock dynamics have been presented in this paper. It is still unknown if these are the only events of innovative qualitative properties in the various regions of this universal relative dynamics algorithm's parameter space. As in astrophysics, developments in the theoretical front coupled with developments in researchers' observational capability (through improvements in computing processes and capacity) guarantee the discovery of new phenomena to improve our understanding of iterative spatial dynamics.

Particular emphasis was placed on the fundamental discrete dynamics bifurcation, equivalent to Hopf's bifurcation in continuous dynamics. Above all, the accent must be on the chaotic behavior of these dynamics, as socio-spatial systems are likely to be found in such states.

ACKNOWLEDGEMENTS

Funding from the National Science Foundation is gratefully acknowledged. Professor Dendrinos would like to thank his research assisant, Richard Desko, for the computer work.

REFERENCES

Dendrinos, D.S. and Sonis, M., Turbulence and Socio-Spatial Dynamics: Toward a Structural Theory of Social Systems Evolution, 1987.

Feigenbaum, M.J., "Quantitative Universality for a Class of Nonlinear Transformations", Journal of Statistical Physics, Vol. 19, 1978, pp. 25-52.

Haken, H., Synergetics, 3rd ed., Springer-Verlag, New York-Heidelberg, 1983.

Lorenz, E.N., "Deterministic Nonperiodic Flow", Journal of the Atmospheric Sciences, Vol. 20, 1963, pp. 130-141.

May, R.M., "Simple Mathematical Models with very Complicated Dynamics", Nature, Vol. 261, 1976, pp. 259-467.

Saaty, T.L., Modern Nonlinear Equations, McGraw Hill, New York, 1967.

Sonis, M. and Dendrinos, D.S., 1984, "A Discrete Relative Growth Model: Switching, Role Reversal and Turbulence", Paper presented at the Sixth Advanced Summer Institute in Regional Science, University of Bamberg, FRG, August, 1984. Forthcoming in Conference Proceedings.

Ulam, S.M., "Experimental Mathematics: the Role of Computation in Nonlinear Science", Communications of the ACM, Special Section on Computing in Theoretical Physics, Vol. 28, No. 4, 1985, pp. 374-384.

Chapter 16

Adoption and Diffusion of Innovations and the Evolution of Spatial Systems

H. BLOMMESTEIN and P. NIJKAMP

1. INTRODUCTION

Technology and economic dynamics are closely intertwined phenomena, and therefore it is no surprise that current economic stagnation has led to an increased interest in innovation as one of the driving forces for structural change (see, for instance, Kleinknecht, 1986). The role of innovation has become central to current economic research; witness the great number of debates on the validity of concepts like 'depression trigger', 'demand pull', and 'technology push'. In this context, Stoneman (1983) has divided this research area into the following parts: the generation of new technology, the diffusion pattern of new technology (including the adoption of innovations), and the socio-economic impacts of these processes. These three elements will be discussed briefly.

First, the way in which new technologies and innovations are induced has been studied quite extensively in the recent literature, in the context of both the 'long waves' discussion, and product cycle theory. Also, the spatial framework of technological innovation has received at great deal of attention, inter alia, in the field of the 'urban incubator' hypothesis (for an extensive review, see Davelaar and Nijkamp, 1986).

The second element, viz. the dispersion of technological innovations, has also received much attention in the past years (see, for example, Brown, 1981), following earlier studies made by Hagerstrand (1967) among others. Despite path-breaking work in this field, the behavioural and quantitative background of many contributions to innovation diffusion has not always been impressive, one of the main reasons being that in several cases a behavioural (especially a micro-economic based) choice theory was lacking. An interesting exception in this field is recent work by Sonis (1986 and Chapter 8) on the relationship between innovation diffusion and spatial change in the framework of ecological dynamics.

Our paper will make a modest attempt to review and extend some essential elements of structural spatial changes caused by technology diffusion and adoption. The focus of this paper will mainly be on the spatial aspects of the diffusion of technological innovations (interpreted here as the design, construction and successful introduction of new - or improved - commodities, services, production processes or distribution processes; see Dieperink and Nijkamp, 1986). In a broader and more comprehensive context, such innovations may occur as clusters, coined 'technological regimes' by Winter (1984). An attempt will be made here to specify a stochastic model for the adoption of innovations and the associated spatial developments.

The third component of technology research, the socio-economic impacts, will be dealt with in a less elaborate manner here. For a more elaborate treatment of this subject, especially concerning relationships between technological change, employment and spatial dynamics, the reader is referred to Nijkamp (1986).

The paper is organized as follows. Section 2 contains the initial version of a stochastic model for the acceptance of innovations in different sectors and different cities, based on the technological potential for adopting innovations. In the subsequent section, an introductory exposition of non-linear dynamic models of the Verhulst type is given. This is followed by an application of the Verhulst type of model, in which the impacts of new activities on the urban economies (in different sectors) are related to competition (in terms of differential attractiveness factors) between cities and sectors in a spatial system. Section 4 provides a link between the acceptance of innovations and their impact on the spatial evolution of the multi-sector, multi-city system. The paper concludes with some reflections on the operational mechanism of such a dynamic model.

2. A STOCHASTIC MODEL FOR ADOPTING INNOVATIONS

The model developed in this paper will be based on a stochastic theory of spatial evolution. An attempt will be made to construct a simple multi-sector, multi-region (or multi-city) model, which is able to describe the impacts of technological innovations in one or more specific sectors of the spatial system concerned. Thus the main focus of this paper will be on the interaction between the evolution of a spatial system and the diffusion of innovation.

Let us assume the following stochastic binary choice model for the adoption of a certain technological innovation (or a technological regime):

$$p(x_i, k, t) = r^k_{it}/(r^k_{it} + h^k_{it}), \tag{2.1}$$

where

$p(x_i,k,t)$ = probability that a city i with size x_i will adopt the innovation in sector k at time t

x_i = size of city i (e.g. measured in terms of population)

r^k_{it} = volume of activities of sector k in the city of size x_i, which are able to generate and implement a certain innovation in period t

h^k_{it} = volume of (remaining) activities of sector k which - for technical reasons - are unable to create and implement a certain innovation.

Clearly, $(r^k_{it} + h^k_{it})$ represents the total volume of activities of sector k in a city of size x_i at time t. It is assumed here that h^k_{it} is independent of city size i, as this limitation is caused by technical reasons, specific for sector k; hence:

$$h^k_{it} = h^k_t. \tag{2.2}$$

Here it is assumed that the volume of activities r^k_{it} can be decomposed into 2 elements, viz. the city size x_i and an innovation acceptance coefficient g^k_t , i.e.,

$$r^k_{it} = g^k_t x_i, \tag{2.3}$$

where g^k_t reflects the fraction of activities of sector k in city i which can technically implement the innovation at hand in this city in period t. The foregoing model is assumed to have the following properties:

(1) $x_i, g^k_t, h^k_t \geqq 0$ (2.4)

$0 \leqq p(x_i, k, t) \leqq 1.$

(2) Technical progress implies the following conditions:

$$\dot{g}^k_t > 0; \dot{h}^k_t > 0; \dot{\kappa}^k_t < 0 \tag{2.5}$$

where κ^k_t ($\kappa^k_t \geqq 0$) is defined as follows:

$$\kappa^k_t = h^k_t/g^k_t. \tag{2.6}$$

In this paper it is assumed that long-run technological progress (i.e. when $t \to \infty$) implies that $\lim_{t\to\infty} \kappa^k_t = 0$. A reasonable specification fulfilling this condition for the time trajectory of κ^k_t , may be an exponential function:

$$\kappa^k_t = a^k_0 \exp(-a^k_t t). \tag{2.7}$$

Clearly, alternative specifications are also possible.

(3) $\lim_{x_i \to \infty} p(x_i, k, t) = 1.$ (2.8)

This condition is in agreement with a hierarchical (rank-size) system. It is easily seen that condition (2.8) also holds if $h^k{}_t = 0$.

(4) $\lim_{x_i \to 0} p(x_i, k, t) = 0$ (2.9)

It is also easy to verify that condition (2.9) holds if $g^k{}_t = 0$.

The foregoing properties imply that p(.) satisfies the additivity condition and hence may be interpreted as a choice probability, as it is easily seen from the binary choice model (2.1) that:

$$p(.) + [1-p(.)] = 1. \quad (2.10)$$

Furthermore, it is also worth noting that permanent long-term technological progress implies:

$$\lim_{t \to \infty} p(.) = 1. \quad (2.11)$$

For further expositions on the shape of the adoption curve of innovations the reader is referred to Allen et al (1978).

Clearly, the degree of acceptance of innovations in a specific city depends both on city size and technological progress. In the light of these observations, it is interesting to question how the dynamics of the spatial system affect innovation diffusion and vice versa. Therefore, in the next section a simple model for spatial dynamics based on a Verhulst dynamic model will be developed. Despite its simplicity, the qualitative properties of our model will be shown to be fairly intricate. Only simulation experiments on a computer are then able to reveal the <u>full</u> flavour of such a space-time model. Nevertheless, some <u>basic</u> qualitative properties shall be outlined in section 4.

3. A SIMPLE MODEL FOR SPATIAL DYNAMICS

In recent years, a wide variety of dynamic (often non-linear) models has been developed in order to describe the impact of a significant exogenous stimulus (e.g. an innovation) on the equilibrium pattern of a dynamic system. A usual specification of general dynamic non-linear models is the Verhulst equation of logistic growth (see Maynard-Smith, 1974):

$$\dot{x} = \alpha x\,(N-x) - \beta x, \qquad (3.1)$$

where α and β are constant parameters, and where N is related to a capacity level (or saturation level) for the systems variable x. The variable x may represent, for instance, the economic performance of the existing system.

Now consider the introduction of a significant technological innovation (occurring in a clustered manner, e.g.: see Mensch, 1979). This new set of activities may be denoted by y; it has an impact on the existing economy as follows (see Batten, 1982):

$$\dot{x} = \alpha x\,(N-x-\gamma y) - \beta x, \qquad (3.2)$$

where y may exhibit the same dynamic pattern as x, so that a nested dynamic process may emerge. This variation through innovations (see Nicolis and Prigogine, 1977) evokes the problem of steady-state solutions (see also Allen, 1976). In recent years this problem has been studied quite extensively in the literature, and further contributions to the field of such multi-actor Volterra-Lotka and predator-prey type models can be found, among others, in Brouwer and Nijkamp (1985), Casti (1982), Dendrinos and Mullally (1984), Dendrinos and Sonis (1984), Pimm (1982), Ralston (1977), and Samuelson (1971).

The strength of those models is that they are able to generate a great diversity of complex dynamic behaviour while retaining simplicity in model structure, although it is still an apparent drawback that most of the models are lacking a testable micro-based behavioural foundation. In this context, an interesting neoclassical approach to the choice process underlying innovation diffusion can be found in Soete and Turner (1984), who used Nelson and Winter's (1982) evolutionary theory of economic growth to analyze the micro-economic level of the adoption and diffusion of new technologies. Their analysis is, however, based on a deterministic approach in which spatial dynamics is not explicitly taken into account; new advances based on stochastic (disaggregate) utility theory can be found in Haag and Weidlich (1984), who tried to develop a probabilistic evolutionary spatial model. A review of such non-linear dynamic modeling efforts can be found in Barentsen and Nijkamp (1986).

In summary, it can be concluded that there is an increasing tendency toward constructing discrete choice models based on a stochastic acceptance (and diffusion) of innovations.

In the present section an illustrative model based on a Verhulst specification will be used as a framework for treating urban dynamics in a spatial system. The fundamental growth equation for city i is supposed to be:

$$\dot{x}_i = \alpha x_i (N + \sum_k \varepsilon^k v^k_i - x_i) - \beta x_i , \tag{3.3}$$

where α is the birth rate of urban activities, β the death rate of existing activities, N the initial carrying capacity for economic activities of the city, v^k_i the volume of new activities in sector k generated in city i (measured in appropriate units), and ε^k the impact of new activities in sector k on the growth of city i.

Thus the expression $\sum_k \varepsilon^k v^k_i$ indicates the capacity expansion augmenting N, due to the introduction and implementation of new activities k.

Next, the growth of these new activities in sector k in city i may be represented as follows (see also Allen et al., 1978):

$$\dot{v}^k_i = \eta v^k_i (e^k_i - \delta^k v^k_i), \tag{3.4}$$

where η is the growth rate of these new activities, e^k_i the volume of employment (or, in general terms, production factors) which might potentially be generated in sector k (i.e., a ceiling for new urban activities), and δ^k a market threshold coefficient in sector k.

In addition, one may assume:

$$e^k_i = \mu^k d^k_i, \tag{3.5}$$

where d^k_i is the demand for the products generated by sector k in city i, and μ^k a (constant) parameter linking the effective demand for k to their potential employment opportunities (usually, $\mu^k \geqq 1$).

Besides, the total demand in city i generated by residents of other cities j, in the absence of spatial competition is equal to:

$$d^k_{ij} = \lambda^k \; x_j \,/\, (p^k_{ij})^{\upsilon} \tag{3.6}$$

where p^k_{ij} is the c.i.f. price of a unit of products from sector k, produced in i and shipped to residents in j; λ^k and υ are normal reaction parameters.

Now the price p^k_{ij} is supposed to depend on communication costs between cities i and j as follows:

$$p^k_{ij} = p^k_i + \phi^k d_{ij}\,, \tag{3.7}$$

where p^k_i is the f.o.b. price, d_{ij} the distance between i and j, and ϕ^k the unit communication cost.

Next, one may introduce spatial competition between cities on the basis of an attractiveness indicator a^k_{ij}, for sector k, which incorporates urban facilities and price levels of sector k

$$a^k_{ij} = \rho \; n_i \,/\, (p^k_{ij})^{\Phi} \tag{3.8}$$

where a^k_{ij} is the relative attractiveness of city i for residents of city j, n_i the share of facilities in city i, while ϕ and ρ are the standardization parameters.

Consequently, the demand inside city i, generated by households outside city i, is co-determined by the relative attractiveness of city i, so that equation (3.6) may be adjusted as follows:

$$d^k_{ij} = \lambda^k x_j a_{ij} \ / \ (p^k_{ij})^{\upsilon} = \lambda^k x_j \rho n_i \ / \ (p^k_{ij})^{\upsilon + \theta}. \tag{3.9}$$

It can easily be seen that the total sectoral demand is calculated directly from (3.9), while (dis)economies of scale may also be incorporated. By substituting (3.9) into (3.5), followed by a substitution of (3.5) into (3.4), equations (3.3) and (3.4) now describe a highly non-linear dynamic evolution of a spatial system composed of competing regions, which might lead to competitive exclusion (see also Johansson and Nijkamp, 1986). Thus, various types of dynamic behaviour may emerge, depending on the initial conditions and the various parameters of the system. As the analytical properties of this model are hard to trace, simulation experiments usually have to be carried out in order to study the stability and equilibrium properties of such a model.

Having now presented a model for spatial competitive dynamics, in the next section we shall integrate the elements of the innovation diffusion model discussed in section 2.

4. INNOVATION DIFFUSION AND SPATIAL DYNAMICS

As mentioned in section 2, technological progress implies that the ratio of activities which cannot technically implement a certain innovation with respect to those which are actually able to do so, is declining (see conditions (2.5) and (2.6)). Clearly, an innovation will only be successfully introduced if it creates a decrease in production costs c^k_{it} ; i.e., if

$$\dot{\kappa}^k_{ij} < 0 \rightarrow \dot{c}^k_{it} < 0 \tag{4.1}$$

Clearly, cost savings will lead to a reduction in the f.o.b. price p^k_i in equation (3.7), so that then sector k in city i improves its competitive position, i.e.

$$\dot{p}^k_i = f(\dot{c}_{it}). \tag{4.2}$$

Consequently, once condition (4.1) is fulfilled, the basic model linking acceptance of innovations to spatial dynamics is composed of equations (2.1), (3.3) and (3.4) (after substitution of the relevant equations).

Now the mechanism of this model and its features may be explained as follows. Suppose a major technological innovation takes place in a cluster-wise manner and penetrates the majority of all sectors k (informatics for example). The spatial-economic spread effects of such an information wave can then be traced as follows. First, there is an initial diffusion of innovation according to equation (2.1). If city i is large, it will probably incorporate a large share of the innovation directly (reflected by a high value of p(.)), while the (hierarchical) spatial diffusion of the innovation concerned will depend on the size of cities in the spatial system.

Next, the time path (i.e., the adoption rate over time) of the innovation depends on the value of κ^k_{it} and its impact on production costs. Clearly, the combination of both processes may lead to well-known space-time processes in dynamic geographical systems (see Griffith and Lea, 1984). The way these combined processes affect the spatial system can now be described in a stepwise way.

(1) Define the existing spatial system with cities i and sector k by means of the above-mentioned state variables and related parameters.

(2) Identify the rate of potential technology acceptance parameters κ^k_{it} for the successive time periods, and calculate the corresponding probabilities p(.) for each city i and each sector k.

(3) If condition (4.1) is fulfilled, one must use (4.2) (as well as the remaining equa-

tions) to analyze the impact of a major innovation (accepted in many sectors) on the dynamic evolution of a competitive spatial system. If, for instance, city size x_i increases (see (3.3)), then p(.) will increase (see (2.1)); a higher adoption rate of innovation will decrease production costs, and hence, the competitive position of city i. Employment growth will take place, leading to a growth in i, and so on.

The dynamic development of such a spatial system might be explored by means of simulation experiments. To some extent, the diffusion mechanism of this model is aligned to the Christaller framework, because city size plays a major role in the adoption rate of innovations. However, because of the distance decay function for communication costs, the model also exhibits a distance-related diffusion pattern. Clearly, technologcial innovation might also lead to a reduction in communication costs. Given the positive impact of large cities on the acceptance rates of innovation, there is some reason to expect that large cities will become larger in our dynamic system. Consequently, beyond a certain threshold level of city size, it might be important to include a negative externalities factor in order to allow the model to generate a broad spectrum of different spatial behaviour (cf. Day, 1982).

5. CONCLUSION

The approach presented in the previous sections was essentially based on the competitive aspects of spatial dynamics. A proper choice and implementation of new technology in a certain place enhances its efficiency, and hence its relative growth chances in a spatial system. This growth was assumed to be caused by a simultaneous occurrence of both producer behavior and consumer behaviour in the adoption of technological innovations.

It should be added that this model, as such, is not fully operational (in terms of empirically based quantitative models). By means of simulation experiments, it may reveal a diversity of space-time patterns, emanating from the interplay of economic and technological key forces (for example, by means of an evolutionary event history analysis).

Our approach emphasizes the impact of technology on production processes and, hence, on the growth of cities through multiplier processes emerging from innovations and related agglomeration forces. In an analogous manner, our model might be used to trace the impacts of new technology on city size in a technology-driven spatial system, through an analysis of agglomeration forces associated with different kinds of technology. Altogether, the potential of this model can be explored further under different economic-technological regimes.

Finally, it is worth noting that a space-time model (like the one considered in this paper), is fairly complex in the sense that it is not possible to represent the content of the model in terms of a few easily tractable qualitative (or quantitative) properties. This is the trade off to be faced if one wants to model the interplay of economic and technical forces in an evolutionary context, whereby a diversity of space-time patterns can be addressed.

REFERENCES

Allen, P.M., Evolution, Dynamics and Stability, Proceedings of the National Academy of Sciences, vol. 73, 1976, pp. 665-668.

Allen, P.M., Deneubourg, J.L., Sanglier, M., Boon, F. and de Palma, A., The Dynamics of Urban Evolution, Vol. 1: Interurban Evolution, Final Report to the U.S. Department of Transportation, Washington D.S., 1978.

Barentsen, W., and Nijkamp, P., Nonlinear Dynamic Modeling in Space and Time, Research Memorandum, Dept. of Economics, Free University, Amsterdam, 1986.

Batten, D.F., On the Dynamics of Industrial Evolution, Regional Science and Urban Economics, vol. 12, 1982, pp. 449-462.

Brouwer, F. and Nijkamp, P., Qualitative Structure Analysis of Complex Systems, in Nijkamp, P., Leitner, H. and Wrigley, N. (eds), Measuring the Unmeasurable, Martinus Nijhoff, Dordrecht, The Netherlands, 1985, pp. 509-532

Brown, L.A., Innovation Diffusion, Methuen, London, 1981.

Casti, J., Simple Models, Catastrophes and Cycles, Kybernetes, vol. 13, 1984, pp. 213-229.

Davelaar, E.J. and Nijkamp, P., Spatial Dispersion of Technological Innovation: The Incubator Hypothesis, Proceedings Conference on Innovation Diffusion, Venice, 1986.

Day, R., Irregular Growth Cycles, American Economic Review, vol. 72, no. 3, 1982, pp. 406-414.

Dendrinos, D.S. and Mullally, H., Urban Evolution, Oxford University Press, Oxford, 1984.

Dendrinos, D.S. and Sonis, M., Variational Principles and Conservation Conditions in Volterra's Ecology and in Urban Relative Dynamics, Collaborative Paper CP-84-49, International Institute for Applied Systems Analysis, Laxenburg, 1984.

Dieperink, H. and Nijkamp, P., Innovation and Regional Policy in a Spatial Context, Research Memorandum, Dept. of Economics, Free University, Amsterdam, 1986.

Griffith, D.A. and Lea, A. (eds.), Evolving Geographical Systems, Martinus Nijhoff, Dordrecht, The Netherlands, 1984.

Haag, G. and Weidlich, W., Model Construction in Dynamic Spatial Theory, Research Paper, Institute for Theoretical Physics, University of Stuttgart, 1984.

Hagerstrand, T., Innovation Diffusion as a Spatial Process, University of Chicago Press, Chicago, 1967.

Johansson, B. and Nijkamp, P., Analysis of Episodes in Urban Event Histories, in Klaassen, L.H. (ed.), Spatial Cycles, Gower, Aldershot, 1986.

Kleinknecht, A., Innovation Patterns in Crises and Prosperity, Oxford University Press, Oxford, 1986.

Maynard-Smith, J., Models in Ecology, Cambridge University Press, Cambridge, 1974.

Mensch, G., Stalemate in Technology, Ballinger, Cambridge, Mass., 1976.

Nelson, R., and Winter, S., An Evolutionary Theory of Economic Change, Harvard University Press, Cambridge, 1982.

Nicolis, G., and Prigogine, I., Self-organisation in Non-equilibrium Systems, John Wiley, New York, 1977.

Nijkamp, P. (ed.), Technological Change, Employment and Spatial Dynamics, Springer Verlag, Berlin, 1986.

Pimm, S.L., Food Webs, Chapman & Hall, London, 1982.

Ralston, B., A Neoclassical Approach to Urban Systems Diffusion, Environment and Planning A, vol. 10, 1977, pp. 267-273.

Samuelson, P., Generalized Predator-Prey Oscillations in Ecological and Economic Equilibrium, Proceedings of the National Academy of Sciences, vol. 58, 1971, pp. 980-983.

Soete, L., and Turner, R., Technology Diffusion and the Rate of Technical Change, The Economic Journal, vol. 94, 1984, pp. 612-623.

Sonis, M., A Unified Theory of Innovation Diffusion, Dynamic Choice of Alternatives, Ecological Dynamics and Urban/Regional Growth and Decline, Paper presented at a Conference on 'Innovation Diffusion', Venice, March 1986.

Stoneman, P., The Economic Analysis of Technological Change, Oxford University Press, Oxford, 1983.

Winter, S.G., Schumpeterian Competition in Alternative Technological Regimes, Journal of Economic Behavior and Organization, vol. 5, 1984, pp. 287-320.

LIST OF CONTRIBUTORS

Ethan Akin
Mathematics Department
The City College
The City University of New York
137 Street & Convent Avenue
New York, NY 10031
U.S.A.

Åke Andersson
Department of Economics
University of Umeå
S-901 87 Umeå
Sweden

David Batten
Department of Economics
University of Umeå
S-901 87 Umeå
Sweden

Martin Beckmann
Department of Economics
Brown University
Providence Rhode Island 02912
U.S.A.
and
Institute for Applied Mathematics and Statistics
Technical University of Munich
Arcisstr. 21
D-800 München 2
Federal Republic of Germany

Hans Blommestein
Department of International Monetary Affairs
Dutch Treasury
P.O.Box 20201
NL-2000 EE The Hague
The Netherlands

John Casti
Institute for Operations Research
University of Vienna
A-1010 Vienna
Austria

Richard Day
Department of Economics
University of Southern California
University Park
Los Angeles, CA 90089-0152
U.S.A.

Dimitrios Dendrinos
Department of Urban Planning
University of Kansas
Lawrence, Kansas 66045
U.S.A.

Jorg Glombowski
School of Economics, Social Sciences and Law
Tilburg University
P.O.Box 90153
NL-5000le Tilburg
The Netherlands

Günter Haag
Institute for Theoretical Physics
University of Stuttgart
Pfaffenwaldring 57/III
D-7000 Stuttgart 80
Federal Republic of Germany

Börje Johansson
CERUM
University of Umeå
S-901 87 Umeå
Sweden

Michael Krüger
Faculty of Social Sciences
University of Osnabrück
Postfach 4469
D-4500 Osnabrück
Federal Republic of Germany

Alfredo Medio
Department of Economics
University of Venice
Ca' Foscari, Dorsodouro 3246
I-30123 Venezia
Italy

Gerhard Mensch
Weatherhead School of Management
Case Western Reserve University
Cleveland, Ohio 44106
U.S.A.

Peter Nijkamp
Department of Economics
Free University
P.O.Box 7161
NL-1007mc Amsterdam
The Netherlands

Tõnu Puu
Department of Economics
University of Umeå
S-901 87 Umeå
Sweden

Gerald Silverberg
Institute for Social Research
University of Stuttgart
Postfach 560
D-7000 Stuttgart
Federal Republic of Germany

Michael Sonis
Department of Geography
Bar-Ilan University
Ramat-Gan, 52-100
Israel

Wolfgang Weidlich
Institute for Theoretical Physics
University of Stuttgart
Pfaffenwalding 57/III
D-7000 Stuttgart 80
Federal Republic of Germany

This series reports new developments in mathematical economics, economic theory, econometrics, operations research, and mathematical systems, research and teaching – quickly, informally and at a high level. The type of material considered for publication includes:

1. Preliminary drafts of original papers and monographs
2. Lectures on a new field or presentations of a new angle in a classical field
3. Seminar work-outs
4. Reports of meetings, provided they are
   a) of exceptional interest and
   b) devoted to a single topic.

Texts which are out of print but still in demand may also be considered if they fall within these categories.

The timeliness of a manuscript is more important than its form, which may be unfinished or tentative. Thus, in some instances, proofs may be merely outlined and results presented which have been or will later be published elsewhere. If possible, a subject index should be included. Publication of Lecture Notes is intended as a service to the international scientific community, in that a commercial publisher, Springer-Verlag, can offer a wide distribution of documents which would otherwise have a restricted readership. Once published and copyrighted, they can be documented in the scientific literature.

Manuscripts

Manuscripts should be no less than 100 and preferably no more than 500 pages in length. On request, the publisher will supply special paper with the typing area outlined and essentials for the preparation of camera-ready manuscripts. Manuscripts should be sent directly to Springer-Verlag Heidelberg or Springer-Verlag New York.

Springer-Verlag, Heidelberger Platz 3, D-1000 Berlin 33
Springer-Verlag, Tiergartenstraße 17, D-6900 Heidelberg 1
Springer-Verlag, 175 Fifth Avenue, New York, NY 10010/USA
Springer-Verlag, 37-3, Hongo 3-chome, Bunkyo-ku, Tokyo 113, Japan

---

ISBN 3-540-18183-0
ISBN 0-387-18183-0