MEMORANDUM FOR: Director of onomic Research

STAT

Attached are some papers left with me by [redacted]. I would like some of your methodologies people to read this and to advise Ed and me of their views on [redacted] evaluation of the Forrester Model.

STAT

Please return attached [redacted]pers.

STAT

Paul V. Walsh
ADDI

Attachments

Date 31 July 1975

FORM 5-75 101 USE PREVIOUS EDITIONS

EWP

File

1 August 1975

MEMORANDUM FOR: Associate Deputy Director for Intelligence

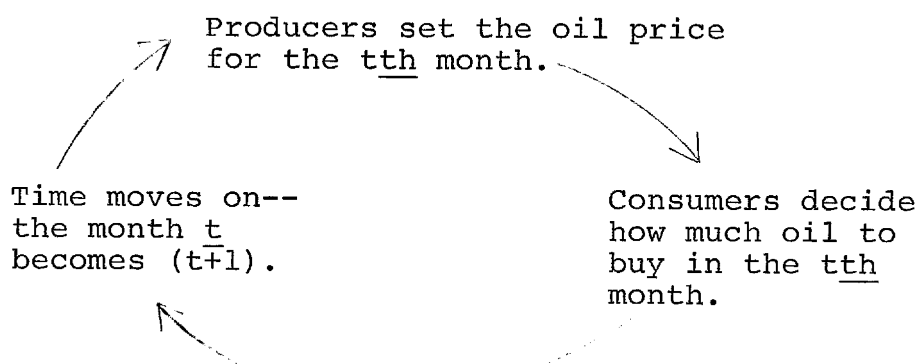
VIA : Director of Economic Research

SUBJECT : An Evaluation of Jay Forrester's "Systems Dynamics" Methodology

STAT
STAT

1. After having read Forrester's books, listened to [redacted] sent three OER analysts to [redacted] three-week course on systems dynamics, and employed an analyst who studied under Dennis Meadows and Forrester, I conclude that Forrester's method is absolutely nothing new. Economists have been developing the same kinds of models for the last twenty-five years. This view is quite heretical to the devout systems dynamicists, who for the last decade have preached the revolutionary nature of their approach, and who still joust quixotically against the imaginary evils of economic models.

2. To see why systems dynamics is nothing new, the simplest approach is to look at the mathematical structure of Forrester's models. All the models are essentially systems of difference equations. For example, suppose that we want to predict international oil prices and production rates for the next five years. In this case, the prices are set by oil-producing countries, who base their decisions on past actions by oil consumers. The consumers, in turn, take prices as given and adjust consumption rates. A model of this process--as depicted by a Forrester snake diagram--could be:



3. In terms of difference equations, this model might be written as:

$$\text{Price}(t+1) = \alpha + \beta\text{Quantity}(t)$$

$$\text{Quantity}(t) = \gamma + \delta\text{Price}(t).$$

_If_ we know the parameters α, β, γ, and δ, then we can _solve the model_ for time paths of price and quantity. As [redacted] says, anybody can do it. Moreover, anyone who has studied economics can recognize the striking similarity between this model and good old supply-and-demand analysis.

STAT

4. Unfortunately, however, we do not know the parameters. We do not even know the functional forms of the relations--on _a priori_ grounds, we could just as well specify the first equation as $\text{Price}(t+1) = \alpha\{\text{Quantity}(t)\}^{\beta}$.

5. Depending on the functional forms and the parameters we assume, a systems dynamics model can yield vastly different predictions. To pick the "right" model, Forrester's disciples suggest that we should first make a reasonable guess, and then test the sensitivity of the model's results to variations in this guess.

6. This approach neglects an important computational consideration. In particular, Forrester's models have dozens of parameters. Should we formulate a relatively small model with only twenty parameters and twenty plausible values for each parameter, then our sensitivity analysis would have to deal with $(20)^{20}$ possible parameter combinations. This would make our analysis of oil markets rather complicated, because the electricity required for a computer to run the model would become a major component of oil demand.

7. In summary, the major difficulty in constructing good predictive models is what it has always been--the problem of specifying the relations in the models. In this regard, polemics about the wonders of systems dynamics are distinctly unhelpful.

Approved For Release 2005/11/23 : CIA-RDP80B01495R000600180019-0

8. Why, then, has Forrester attracted so much attention? My guess is that he is brilliant in popularizing difference equations. A difference-equation model, when written as a set of equations, looks like just another complicated mass of mathematics. But when the same model is described by a snake diagram, and when the model's results are graphed as time paths, then the model becomes a dramatic analytical edifice. Since this is what people want, and since Forrester has fervently met the demand, he is famous.

9. In response to the Forrester cult, I think that the DDI should take a positive attitude. Those who advocate systems dynamics models should be told that the Agency certainly wants to promote sound analytical inferences. To the extent that the Forrester disciples can justify the relations in their models, and to the extent the models address important policy questions, the contributions of systems dynamics will be most welcome.

10. Nevertheless, Agency analysts should not be forced to study Forrester's polemics. His books are available for those who care to read them, and OER personnel will be happy to answer any technical questions that the books omit.

STAT

Chief
Systems Development Group
Office of Economic Research

28 July 1975

MEMORANDUM FOR THE RECORD

SUBJECT: Summer Computer Simulation Conference

The Summer Computer Simulation Conference was held at the St. Francis Hotel in San Francisco on 21-23 July. Approximately 200 of the 450 attendees presented papers at the conference--about 10 percent of the papers were prepared by foreign authors.

The theme for the conference was the same as the conference theme of two years ago: Simulation Credibility. Session topics at the conference covered a wide range: computer hardware, software, simulation theory, simulation applications in the fields of energy, chemical engineering, medicine, transportation, etc. Similarly, the spectrum of the papers presented in the sessions was also broad. In the session which I chaired, Social and Political Models, the models dealt with governments as information processing systems, public budgets, the structure and dynamics of sociological systems, and secure environments. At the last minute, the authors of one paper for the Social and Political Model session were unable to attend. Fortunately, [redacted] of the Office of Political Research was able to present an excellent paper, [redacted] Because the Proceedings of the conference were printed prior to the conference so that complete papers would be available to the attendees for reference, [redacted] paper will not be published. Separate copies were given to those requesting his paper at the conclusion of the session.

A highlight of the conference was the first (I believe) major presentation on the status of the MIT National Socio-Economic Model. The papers were presented in a special session on Tuesday evening by the members of Forrester's group at the Sloan School of Management. The papers dealt with the objectives and the overall status of the project, the dynamics of economic fluctuations, the production and work force sectors of the model, and certain issues relating to the utility and credibility of statistical modeling techniques frequently used, for example, in econometrics. These presentations, together with the papers published in the Proceedings, strongly suggest that the work of Forrester's group will have a great and lasting effect on all facets of social and political analysis.

The work is too comprehensive and too complex to be detailed here, but there are several important facts and personal judgments which I believe are especially significant:

(a) With respect to status, after spending a full year in carefully working out the broad design of the model (Phase I), efforts since September 1974 have been focused on the development of the model's six kinds of sectors: production sectors which include consumable, durable goods, soft goods, capital equipment, building construction, agriculture, natural resources, energy, etc.; demographic sectors; household sectors replicating economic categories--labor, professional, unemployed, welfare, retired; financial sectors; government sectors; consumption sectors; and a foreign sector to treat trade and international monetary flows. The second phase of the project is nearing completion. All sectors are operational, and critiques of their behavior are being solicited from experts in each field--a massive interdisciplinary approach to model development. Late in 1975, and thereafter, the model is expected to produce analyses of national policy issues and alternatives, thereby beginning the fulfillment of its objectives.

(b) Forrester is going to great lengths to use the model to elicit and exploit expert opinion and criticism prior to any publication of "results"--no doubt to preclude, or at least to minimize, the considerable furor that followed the publication of Urban Dynamics and World Dynamics.

(c) Although focusing on the United States, the model is generic, i.e., the structure of the model is designed to be applicable to any region, country, or combination of countries, by determining and introducing the parameter coefficients which characterize the area being modeled.

(d) Concurrent with the development of the model--it seems to me--Forrester has conducted a companion effort focused on current methodologies relating to his work. Reflecting again a facet of the debates following his publications, i.e., Why is your methodology superior?; What's wrong with what we're doing?, he has begun to spell out and document the answers to these questions in detail. This is especially apparent in the case of economics. Peter Senge of Forrester's group has written a paper which goes to the guts of econometrics--the statistical methods used in model building--and he asserts that these assumptions are not valid under realistic conditions.

It is not unlikely that Forrester's current work will revolutionize analysis--and, probably, much of the intelligence process. Regardless of their shortcomings and related misunderstandings, World Dynamics and Limits to Growth have clearly changed man's perspectives of the future and time scales on a global basis. I expect this work to have an even greater impact on methodology. The reasons for this are straightforward: Forrester's methodology readily incorporates directly, and exploits, subjective expert judgment and experience as well as empirical facts and evidence; it greatly facilitates communication among experts of different disciplines and between them and the non-experts--the policy makers and the general public; and he has explicitly identified fatal flaws in some current analytical methods--especially econometrics. Importantly, Forrester's method is relatively simple and straightforward: it is not dominated by mathematics and statistics--practically anyone can "do it." In sum, although long-suffering and unappreciated, Forrester seems to be well along in building a methodological nuclear weapon while concurrently reducing some others which are now widely used from TNT to something less than dynamite, e.g., demonstrating that conventional economic methods may not be merely dismal but, unfortunately, dysfunctional.

From the papers published to date--or Forrester's presentation at the World Future Society meeting in June, or Peter Senge's presentation at the Summer Computer Simulation Conference--my evaluation of the recent work of Forrester's group may seem overdrawn. Forrester spoke directly to the implication of Senge's work when I talked privately with him at the World Future Society meeting. At the conclusion of the presentations on Tuesday evening, I told Senge that Forrester had sent me the December 1974 version of his paper and spelled out my interpretations of its conclusions. Senge concurred in these conclusions.

Other items:

(a) Attendance at the regular sessions I attended generally numbered 20 to 30. The attendance at the Social and Political Model session--a session which had not been a part of earlier conferences--varied between 35 and 45.

(b) Dr. Albert Stone with whom I had worked while at the Applied Physics Laboratory of the Johns Hopkins University, also attended the conference. He will be in charge of organizing all sessions on energy for next year's conference which will be held in Washington, D.C. I agreed to recommend a CIA officer working in this area to chair one of the energy sessions.

(c) The fact that [redacted] and I openly participated as CIA officers appeared to generate no problems or curiosity whatsoever, quite the contrary. STAT

[redacted] STAT

C/ISTB/FTD

UNDERSTANDING SOCIAL AND ECONOMIC CHANGE IN THE UNITED STATES

Jay W. Forrester
Germeshausen Professor
Massachusetts Institute of Technology

I. INTRODUCTION

The MIT System Dynamics Group is currently well advanced on a national model of social and economic behavior.* It is a system dynamics model and so is very different from the more common econometric models. The present controversies about the economy, uncertainties about the causes of inflation, and debates about economic theory all suggest the need for a new approach to economic dynamics. We believe there is an excellent chance that a comprehensive system dynamics model incorporating the structures that generate economic fluctuations, growth, and environmental restraints can complement other approaches and can fill in where other methods of analysis have been unable to answer important questions.

The system dynamics model now nearing completion should yield substantial new understanding of the major social and economic pressures confronting the United States. Within the next few months the model will be far enough along for us to start examining the forces underlying inflation, the nature of the new economic mode that can simultaneously produce inflation and unemployment, the impact on standard of living as the United States buys more energy and resources from abroad, the consequences of various methods of recycling money paid for oil imports, the effect on exchange rates from foreign manufacturing by multi-national corporations, and the economic forces arising to reverse the historical flow of people from agriculture and manufacturing toward government services.

The inadequacy of past precedents and rules of thumb is becoming evident. Only through modeling will we be able to develop more promising policies for the future, and to communicate new insights widely enough to establish public support for measures necessary to cope with an increasingly difficult future.

II. SYSTEM DYNAMICS

The economic model now being assembled is a system dynamics model. As such it differs from more traditional economic models in structure, sources of input information, nature of validity testing, and purpose. A brief description of the system dynamics approach should help in understanding the model.

System dynamics is a way of combining personal experience with computer simulation to yield a better understanding of social systems. The field of system dynamics has been under development at MIT and elsewhere since 1956. Twenty or more books have been published on subjects ranging from corporate policy to major world interactions. The principal books from the research programs have each been adopted as texts in dozens of universities.

A. Background of System Dynamics

System dynamics is perhaps best described in terms of the background threads on which it builds. In Figure 1, three earlier developments--traditional management of social systems, feedback theory, and computer simulation--combine to become system dynamics. Traditional management is the process used to govern social systems throughout history. Feedback theory or cybernetics is a body of methods and principles developed during the last hundred years dealing with how decisions, and the way they are imbedded in information channels, cause the dynamic behavior of systems. Computer simulation allows one to determine the time-varying behavior implicit in the complex structure of a system.

People start the traditional management process by observing the world about them, noting the pressures and reactions of people and groups, and detecting the linkages and flows of information and influence. From these observations people form mental images of the structure of a social system. From the mental images they attempt to anticipate what will happen next and how a different policy might make the system behave more desirably.

Traditional management processes guide our personal lives, family affairs, cities, countries, and international relations. It is the nearly universal approach to directing human activity. Because it is the basis of civilization and because it has served well, no quick or radical break with past tradition is either possible or desirable.

Traditional management, based on observation and judgement, has great strengths, but it also has serious weaknesses. Any new contribution to better management of social systems must start from the present practices and move gradually toward improvement. Any better method of decision-making must build without discontinuity on the strengths of traditional management while compensating for the weaknesses.

The greatest strength of traditional management comes from the wealth of information available from the separate observations and experiences of people. In the mental stores of knowledge are probably a thousand or million times more information than has been converted to written from in libraries. In turn, written descriptions cover a thousand or million times the scope and richness of information that is available in measured and numerical form. If we are to improve on social decisions, we must be able to build on the most comprehensive information base available--the observations, knowledge, and judgement stored in people's heads. System dynamics uses that descriptive information along with any available written and numerical information.

But traditional management has several serious weaknesses. System dynamics helps to alleviate those weaknesses.

The first weakness in traditional management arises

*The National Modeling project is being supported by the Rockefeller Brothers Fund.

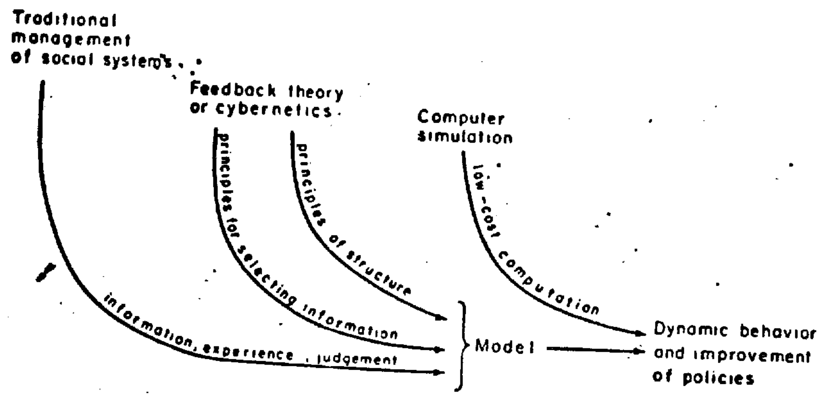


Figure 1. Background of system dynamics.

from the very wealth of information that is the greatest strength. In fact, we have too much information. We are flooded and overwhelmed with information. The traditional processes contain no general principles or organized philosophy for picking the relevant from the extraneous information. As indicated in Figure 1, principles drawn from feedback theory assist in choosing from the excess of information that which is relevant to the behavior modes of interest.

The second weakness of traditional management arises from lack of organizing principles for the structuring of information. Even if the first weakness is overcome and the relevant information and relationships are chosen, no guidelines exist for organizing the chosen assumptions into a structure that explains the observed system behavior. Again, feedback theory offers principles [8] for simplifying and organizing the structure of a system.

But even if information is effectively selected and usefully organized into a relevant model, traditional management encounters a third weakness. Even when assumptions are explicitly stated, the human mind is not well adapted to determining the future time-varying consequences of those assumptions. Different people, even when they accept the same assumptions and structure, often draw contrary conclusions. A consensus is hard to reach, and even a majority opinion may be incorrect. As suggested in Figure 1, computer simulation can be used to determine, without doubt, the future dynamic implications of a specific set of assumptions.

To summarize, system dynamics starts from the practical world of normal economic and political management. It does not begin with abstract theory nor is it restricted to the limited information available in numerical form. Instead it uses the descriptive knowledge of the operating arena about structure, along with available experience about the decision-making. Such inputs are augmented where possible by written description, theory, and numerical data. Feedback theory is used as a guide for selecting and filtering information to yield the structure and numerical values for a computer simulation model. Because the resulting models are too complex for either intuitive or mathematical solution, a computer simulates, or plays the roles, of the many participants in the system to determine how they interact with one another to produce changing patterns of behavior.

B. Necessity for Models

Models are not new in social decision making. System dynamics does not for the first time introduce models into the social and political process. Models have always been the basis for the traditional management methods.

Every decision we make is based on a model. One does not have a family, city, or nation in his head. He has only images, relationships, and abstractions from real life. These perceptions are models in the same sense that the word is used in system dynamics. One uses observation to form a mental image, or model. The mental model becomes the basis for decisions.

System dynamics does not impose models but is a way of improving on the models that would otherwise be used to manage human affairs. The system dynamics model is more explicit than a mental model, so it can be communicated with less ambiguity. A system dynamics model is more carefully structured in accordance with dynamic principles, so it better relates underlying assumptions to system behavior. A system dynamics model can be simulated on a computer, so, unlike a mental model, its behavioral implications can be determined precisely.

C. Comparing System Dynamics with Econometrics

System dynamics is a departure from conventional methodology in economic modeling. Most major economic models have used econometric methods to convert historical numerical time-series data into parameters for an assumed structure of equations. But such methods have failed to answer pressing questions about fundamental behavior arising from social, economic, and environmental interactions. The current national issues are so critical that a new approach should be tried. Compared to an econometric model, a system dynamics model:

a. makes greater use of descriptive information and managerial and political experience
b. incorporates a broader range of variables and encompasses the many relevant disciplines outside of economics
c. uses numerical data from real life in a different way--in model construction to complement descriptive information, in validation to compare with corresponding output data from the model
d. generates social and economic fluctuations and growth from the internal feedback structure without using exogenous variables to drive change
e. includes important social and psychological variables for which statistical data are not

available.

f. explains how "structural changes" occur as the economy moves into new modes of behavior that are not represented in past time-series data
g. facilitates incorporating the wide range of non-linear structures that generate so much of observed real-world behavior
h. emphasizes the conservation of flows by including the buffer stocks that decouple the instantaneous flow rates
i. distinguishes more sharply between real variables, their money value, and information about them to capture the dynamic interactions between the real, money, and information aspects of a system
j. encourages construction of a deeper substructure of feedback loops to represent causal mechanisms underlying macroeconomic behavior
k. organizes structure so that each parameter has independent real-life meaning in the operating world and can be individually drawn from and checked against descriptive and quantitative information available at the place in the real system to which the parameter applies
l. serves as a more effective communication medium for resolving disagreements because of the way both model structure and parameters correspond with descriptive knowledge in the operating world
m. places more emphasis on the importance of internal structure
n. focuses more on understanding the reasons for observed behavior and on developing policies to produce better behavior, and focuses less on prediction
o. combines over a greater time span the short-term with long-range human objectives
p. permits a wider diversity of contact between the model and the real world to make validation more persuasive.

D. Avoiding the Practical Difficulties with Statistical Models

A system dynamics model can circumvent the major practical shortcomings that keep econometric methods from deducing correct parameter values and from correctly evaluating the validity of hypothesized structures. Dissatisfaction with econometric models is widely reflected in the economics literature. Although failures have usually been attributed to inadequate data, a more fundamental reason is emerging for the deficiencies in econometric models.

The statistical methodologies are based on precise theoretical foundations regarding correspondence between the hypothesized model structure and the structure of the real system, the nature of random disturbances in the real system, the characteristics of auto and cross correlation, the frequency of sampling of collected data, and the absence of measurement errors in the data. As recognized by almost everyone, none of the underlying theoretical requirements are fully met, so the practical application of econometrics rests on the assumption that the theoretical requirements are approached sufficiently closely for the methods to be applicable.

As long as econometric methods are applied only in real situations, no final test of methodological accuracy is possible because the true values of the real system parameters are not available for comparison with parameter values that have been deduced from statistical analysis. In ordinary practice the estimates of validity are entirely internal to the statistical process itself, so the estimate of validity also rests on the assumption that underlying theoretical requirements are adequately met.

But how closely must the demands of theory be approximated? The answer cannot be found in the real-life setting but can be determined in controlled laboratory experiments. A dynamic, feedback model can be defined as the "real" system and used to generate, without exogenous driving variables, time-series data comparable to that collected in real situations. The data-generation model stands in the place of the real system. Separately, an estimation model is used by a statistical analyst as an approximation to the "real" system as is done in the usual data-analysis procedures. All conditions of the process are completely controllable. The statistical analyst can be allowed to know as much or as little about the structure of the "real" system as desired. The random noise processes within the "real" system can be controlled. The interval between data samples can be independent of the dynamic processes within the "real" system. Controllable errors can be inserted in the data collected from the "real" system. The degree of correspondence between the structure of the "real" system and the structure of the estimation model can be controlled. And finally, perfect knowledge is available about the structure and parameters of the "real" system to permit a definitive evaluation of the parameters obtained by the statistical methods.

Such comprehensive examination of the practical aspects of statistical model-building has recently been started. The first tests, done on single-equation least squares regression analysis, reveal the consequences of various deviations from the theoretical assumptions underlying the standard procedures. It is beginning to appear [9] that least squares regression, which is probably the statistical method most often used by social scientists, is highly sensitive to surprisingly small departures from the idealized theoretical foundations. In fact, the statistical methods break down and become misleading with such minor departures from perfection that meeting the theoretical requirements closely enough seems inconceivable.

The laboratory tests indicate that the generalized least squares data analysis can give not only major errors in the estimates of parameters but also misleading indications from the internal validity measures. Accurate parameters can be obtained in a correct structure for the estimation model along with validity measures that suggest low confidence; depending on such results of the analysis would lead one to discard correct parameters and structure. Or, at other times, inaccurate parameters can be obtained in a correct structure of the estimation model along with validity measures that suggest low confidence; the likely action based on the low validity measure would be to discard the correct structure in the estimation model.

These laboratory experiments indicate that statistically-derived parameters are likely to be further in error than the estimates made for system dynamics models from direct observation of processes in the actual social system. Unlike econometric models, system dynamics models are structured to facilitate parameterization by direct observation of the real-life decision-making processes. In a system dynamics model every parameter can have an independent real-life meaning. It can be discussed and evaluated in terms of its own real-life existence. Those familiar with the structure and policies of the particular part of an actual system have the necessary information to evaluate the reasonableness of a parameter value. And reasonable values are usually sufficient because behavior of a typical complex social structure is surprisingly insensitive to most parameter values. [1, pp. 57-59, 105, 118-119.]

III. DYNAMICS TO BE REPRESENTED

The new model of the economy addresses a wide range of dynamic behavior. The structure should be detailed enough to generate most of the major characteristics observed in the real economic system. Dynamic behavior can be described in terms of periodicities, modes, and time horizon. These characteristics are discussed below.

A. Periodicities

One might separate into individual models the different rapidities of response in a system. Each model would be designed to examine a particular kind of behavior. On the other hand, in a national socioeconomic system the wide range of inherent periodicities may overlap and influence one another. Oscillatory modes of similar duration might pull together and entrain one another into a single dynamic behavior. Or they might remain separate but enhance one another's effect. Because so little is known about the diversity of such interactions, the present model combines a wide range of dynamic phenomena into one structure.

The model will simultaneously be able to create a wide span of time responses--from the business cycle on the short end, through intermediate interest-rate-capital-investment cycles, to the once-in-history transition from growth to equilibrium. To do so will require a structure containing the short time constants associated with inventories, backlogs, and bank balances. The model also represents the slower processes associated with accumulation of buildings and machinery, and the movement of people. At the long end of the time spectrum, the model should contain population growth, land occupancy, and resource depletion.

The business cycle of some three to seven years duration probably arises from interaction between inventories and employment. More commonly, the business cycle has been attributed to capital investment, but the planning time and life of capital are long enough to suggest that the primary contribution of capital investment is to a longer fluctuation in the economy. By reaching down to the fine structure of employment, inventory management, and materials procurement, the model should deal correctly with the business cycle [10].

In the intermediate range of behavior, some evidence suggests a cycle (the Kondratyev wave) in the economy of some fifty years duration. The existence of such a long wave is important to explore. If it exists, its last collapse was into the depression of the 1930's, and its next could be imminent. The long wave may have a more powerful effect than the business cycle.

At the slowest end of the time spectrum is the life cycle of economic development. For the first time, the industrialized societies seem to be moving into a new phase of the life cycle of growth. If so, the United States is now at a point of departure from past trends and expectations such as occurs only once in the history of each civilization. As shown in Figure 2, the economy now appears to be in the "transition stage" between growth and some form of future equilibrium in population and industrial activity. In the past, we have experienced exponential growth, with a doubling of economic output every twenty or thirty years. Such growth cannot go on forever. There must be a leveling out or a peaking and decline. The debates are about when and how the past kind of growth will end, rather than about whether or not it eventually must end.

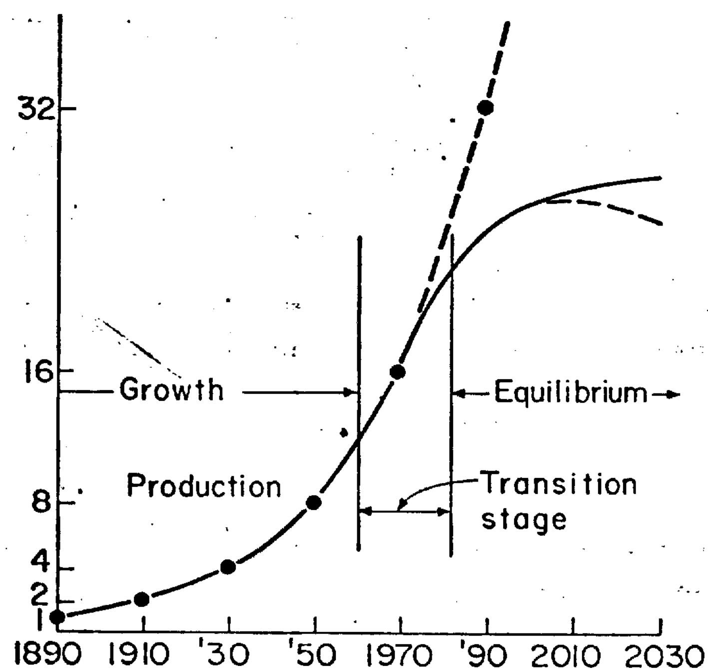


Figure 2. Life cycle of economic growth.

The transition period is about half-way up the life-cycle curve of economic growth and occurs some two doubling times before the economy reaches a peak. Doubling times have been some twenty or thirty years long. At present, about fifty years before we can expect a peaking of population and economic activity, is the time when social and economic forces are pushing us out of the old growth mode.

I believe we are in the transition stage. The transition stage is consistent with the social, environmental, and inflationary forces that are developing. A model that is to cope with today's issues must incorporate not only the recurring processes of past business-cycles, but also the transition process triggered as an economy and its population begin to exploit fully the available energy, resources, agricultural lands, and water.

The time of greatest social and economic stress occurs during the transition stage, not during equilibrium. When equilibrium has been reached, the nature of the new mode of social and economic behavior will be understood and accepted. But in the transition stage, sufficiently great forces arise to overcome the old engines of growth. Laws, attitudes, management methods, traditions, values, expectations, and religions must all change. The transition stage is the time of turbulence as the system moves out of the growth mode.

A socio-economic model to deal with today's questions should encompass the short-term dynamics of the business cycle in concert with the structures that may produce a 50-year long wave and against the background of the life cycle of economic development. For the first time we may face the triple coincidence of a business downturn, a long-wave collapse, and the pressures of the transition region. The three could combine to depress economic activity and the standard of living.

B. Modes

System modes can be described in terms of the associated restraints. The typical economic system is probably unstable in a free mid-range region where restraints are not dominating behavior. In such a restraint-free region there will be a strong tendency for the economic system to move toward one of the possible restraints. If this is correct, shifting modes of economic behavior can be described in terms of the successive restraints that dominate behavior.

At the beginning of industrialization when there was ample land and an excess of labor, the restraint was insufficient capital equipment. The whole philosophy of capitalism arose from the capital shortage that determined the pace of economic development. But after World War II the United States economy moved out of the mode dominated by a shortage of capital.

In the 1950's and 1960's the United States economic system has been characterized more by a shortage of manpower than by a shortage of capital. This has been indicated by chronic shortage of labor and by use of much of the capital plant only 8 hours a day instead of 24. In many businesses capital has been in sufficient supply that it is scheduled for the convenience of people. Because labor has become the principal restraint, labor demands set the style and pace of the economic system.

Now the United States is moving once again into a new mode characterized by shifting restraints. The mode of labor shortage is giving way to the mode of environmental shortage. Environmental shortage exists in terms of space, agricultural land, pollution-dissipation capacity, resources, and energy.

The model should move through these various modes not in response to exogenous driving assumptions, but under the progression of its own internally generated social and economic forces. In effect, the model embodies a theory of economic development that can be tested by seeing if it will generate the modes of behavior that have been observed.

If the model is successful in making the transitions between social and economic modes, it should help reveal the nature of the transitions, their causes, the forces to be expected during any specific change, and the policies that would establish and sustain a desired mode.

C. Time Horizon

To generate the wide range of periodicities and modes, the model must be conceived in terms of a long time horizon. The model should create appropriate growth and fluctuation from the year 1800 to 2100. The model should generate behavior typical of the past as a base from which to anticipate the future. Generally speaking, forces and structures visible at any point in time dominate a very long way into the future. If the fundamental nature of the present system is carefully examined, most essential dynamic mechanisms for the next several decades should be detectable.

A model with a 300-year time horizon imposes special demands on its parameters, variables, and structure. Any constant in the model must be constant for 300 years and must transcend the entire life cycle from early industrial growth through all the traumatic changes of the present and near future and into a mode of equilibrium behavior very different from the past. Therefore, parameters must be extremely fundamental. They become descriptions of human psychology at a stable level that does not shift in response to the immediate political and economic conditions. Any social attitudes, ethical principles, or human preferences that themselves evolve from the surrounding economic and geographic circumstances must be cast as variables responsive to the socio-economic pressures. The model must be anchored on concepts so fundamental that they represent human psychology and descriptions of nature that are not subject to change by the forces the model itself is exploring. Concepts that are more fleeting must be formulated as variables and generated by the forces from which they arise.

IV. SOCIAL AND ECONOMIC ISSUES

A model should be constructed for specific purposes. The purposes come first and shape the design of the model. The purposes for this socio-economic model can be described in terms of the issues to be explored.

A. Inflation

The model is planned for examining the forces underlying inflation. Today's inflation is a much deeper issue than revealed by the public press or by explanations in the economics literature. It is much more than a question of inflation versus unemployment.

Present inflation arises from major imbalances in the economy. Some two-thirds of employment is outside of agriculture and direct production. This constitutes a very high overhead in government, education, and the service industry. Two-thirds of the working population in overhead is probably too great for the economy as we move out of uninhibited economic growth into a period when production is progressively more limited by environmental restraint while, at the same time, population continues to rise. Much of the inflationary pressure comes from governmental efforts to sustain a rising standard of living when real output per capita is running into inherent barriers.

The efforts to hide, by monetary and fiscal means, the fundamental changes now occuring in the industrialized economies are driving inflation. Changing social attitudes, greater complexity arising from crowding, and increasing capital investment required as space and resources become overcommitted are all interlocked in the inflation syndrome.

Actions taken to counter inflation, like escalator clauses in contracts and "indexing" of future payments to compensate for inflation, may accelerate inflation. Such changes in the legal structure of the economy should be studied in a realistic model before being put into practice.

B. Recession, Depression, and Unemployment

Most important in the near future, the model should replicate changes now occurring in the economy. By duplicating the gathering economic stresses, the model should be an effective vehicle for better understanding the causes and cures for recession coupled with inflation and a vulnerable credit and banking system.

Current national economic actions are those that might be appropriate to a normal business-cycle recession. But the forthcoming changes may be far more than a recession. If we are at the peak and entering the steep decline of a 50-year capital-investment cycle, the cause is over-investment in office buildings, automobiles, and many kinds of production facilities. Under such circumstances, national policy to sustain investment may simply delay a needed realignment within the economy.

To the extent that the final slowing of long-term economic growth is behind the present economic stresses, the fundamental issues are more demographic and environmental in nature than economic, and solutions call for redirecting the national focus of attention. A substantial percentage of the work force is devoted to creating growth and to coping with the strains arising from growth. When growth slows in the late part of the economic life cycle, substantial unemployment will develop from jobs that need no longer be filled. The day of reckoning can be delayed but not escaped. The further ahead we recognize changes imposed by the life-cycle of economic growth and work toward an orderly evolution of the socio-economic system, the less traumatic will be the realignments. Because of the extreme complexity of interactions between social, financial, technological, and demographic forces, only a comprehensive model will give access to the behavior we need to understand.

The dynamics of the Great Depression of the 1930's should be examined in search of a better understanding of causes. Was it merely from governmental mismanagement of the financial system? Was it random bad luck? Was World War I significant? Was it the collapse phase of a 50-year cycle? Did it arise as a consequence of the major migration from farm to factory? Can it recur?

C. Wage and Price Controls

The model contains separate price and wage generation in each production sector. By suppressing changes in one or more of these, the effect of price and wage controls can be examined. Prices and delivery delays (availability) influence flows in the model, so the model should realistically respond to controls and should show how controls might transform price pressures into other social and economic pressures.

D. Nature of Economic Growth

The model, by linking population, environment, knowledge generation, and technological contribution to productivity, should provide insights into the nature and future of economic growth. The world has been pursuing economic growth with success in some countries and lack of success in others. The model can be used to examine reasons for past growth and to examine whether or not the gains of the past can be sustained.

Economic growth is inherently a transient process. It cannot continue forever. But where does it lead? Will the higher standard of living be sustained in the future or fall back? The answer depends on how population, technology, and nature interrelate. The standard of living rises when production grows faster than population. But as limitations of energy, resources and space slow the growth of production, growth in population may be slowed more or less quickly with a consequent retention or loss of the standard of living achieved by past economic growth.

The end of economic growth in the equilibrium stage of Figure 2 can take many forms. If population rises faster than production in the late stages of growth, standard of living peaks in the transition stage and then declines. The end point of economic growth can move toward conditions found in India. People have strived mightily to initiate a US-type economic growth in India without success. The reason may be that India has already arrived at the end of the economic growth life-cycle --a condition in which population can exceed the capacity of the country's land and resources.

Economic growth in the United States (and, based on external resources and markets, also for Japan and Western Europe) has been a very special case. Overlooking the way the American Indian was evicted, the United States was a huge empty country of rich agricultural land and plentiful resources and energy. Economic growth has consisted of filling that land with population while using the bounty of nature. But such a growth process is transient. It cannot continue. As growth falters, the nature of the socio-economic system changes. To what? Many choices lie in the future. Now is late but not too late to choose. But we must first have a way to examine the alternatives and the policies leading thereto.

E. Agriculture

As an economy moves through its growth life cycle, the role of agriculture changes. At the beginning the economy is rural. As capital accumulates and labor becomes scarce, and if energy is ample, agriculture becomes capital intensive. The productivity per worker in the field increases but not necessarily the output per unit of energy input. American agriculture is actually a low-efficiency converter of petroleum calories into food calories, a useful process when energy is plentiful but less effective when energy shortages develop. Toward the end of economic growth as labor becomes excessive and energy and land become scarce, a transfer of labor back to agriculture is probably necessary. Such a reversal of labor migration should be examined because it affects government policies on housing, transportation, welfare, and unemployment compensation.

F. Population and Standard of Living

As the capacity of a country becomes fully committed, a tradeoff must exist between population and the standard of living. The higher the population the lower will be the standard of living. The compromise faces each country. Whether or not the issue is recognized will mold the future character of the society. Population versus standard of living does not concern underdeveloped countries alone. Population density of Massachusetts is one and a half times the population density of India. Internal capacity of Massachusetts to support that population is probably little better than India's. Industrialized countries have been buying low-cost resources and selling high-priced manufactured exports. But as the balance shifts and resources become scarce while the capability to manufacture spreads, the status of the have and have not nations begins to converge. As every aspect of the world's capacity becomes more fully committed, each country will begin to face life within the scope of its own land and resources. Major internal realignments will be occurring. The economic mode of the future can be substantially different from the past. A national socio-economic model, if comprehensive enough, should help anticipate the actions necessary in the readjustment period.

G. Education and Economic Change

Education is a form of capital investment. Education increases skill, production output, and human satisfaction. But much of education has been used to fill the inventory of skills; with the inventory filled, only replacement is needed. The educational system, like several other parts of the service sector, shifts from being inadequate to being over-extended as the economy passes the steepest part of its growth. Evidence of excess capacity in higher education is appearing. Governmental action to withstand developing pressures and to sustain historical trends may only lead to more drastic readjustments later. By interrelating consumption demands, productivity, technology, and balance of skills, the model should generate the rising and falling balances between sectors of the economy.

H. Capital Utilization

During growth, production is primarily dependent on capital and labor. As long as land and energy are available, the standard of living rises as the capital-to-labor ratio increases. But resources, energy, and environmental capacity are consumed. In time, growth impinges on natural restraints. When the environment is fully committed, total production is limited by the capacity of nature and the standard of living is determined by the nature-to-population ratio. Under the latter circumstances, the capital-to-labor ratio becomes irrelevant to total production, which is set by environmental limits. Instead, the capital-to-labor ratio becomes a social issue. Capital intensive production with few people working can be combined with income redistribution for supporting others. Or, labor intensive production can be chosen in response to a social decision giving each person a right to a job. Such issues need clarification.

I. Taxes

The consequences of collecting taxes from different points in the economy should be examined. Congress and state legislatures endlessly debate the merits and equity of who to tax and how. What are the relative advantages of property tax, personal income tax, corporate income tax, sales tax, or value-added tax. Such questions may have only a short-term significance. In the longer run, prices and wages can readjust so that money flows to the point from which it is extracted. The structure of the economy suggests that the total tax levied may be far more important than the method. If so, types of taxation may have little leverage for inducing social change, in spite of the rhetoric addressed to taxation issues. A comprehensive model incorporating various channels of taxation and containing the processes of price and wage setting should permit evaluation of tax policies.

J. Balance of Payments

As the prices of energy and resources rise relative to manufactured goods, the balance of payments deficit alters exchange rates and drives down the internal standard of living. Governmental policies affecting trade can have different long-term and short-term effects. Policies to alleviate immediate pressures can accentuate future problems. The trade-offs need to be evaluated in terms of the internal and external economic consequences. A dynamic economic model with an external sector from which to buy resources and sell goods should allow the study of national coupling to the international economy.

K. Fiscal and Monetary Policy

Economic and political debate has centered on policies for managing the economy, enhancing economic growth, and reducing unemployment. But have the policies been effective? The reduced amplitude of business cycles in the post-war years is often cited as evidence for effectiveness of such policies. But the suppressed business cycle may instead reflect other causes--the effect of governmental transfer payments, the labor-shortage mode of the economy, or the rising phase of a Kondratyev long wave caused by labor shifts and the dynamics of capital accumulation. Past policies may have contributed more to inflation than to reducing employment or stabilizing the economy.

Reviewing past policies is important, lest incorrect interpretations lead to misguided future action. With internal monetary and fiscal sectors and with taxation, debt management and government expenditures, the new economic model should offer a basis for resolving debates over Keynesian versus monetarist proposals for government intervention in the economy, and how each is related to growth, stability, and inflation.

V. STRUCTURE OF THE MODEL

The structure of the socio-economic model is intended to be general and to apply to any country having agriculture, consumption, manufacturing, and money. By concentrating first on the United States economy, while keeping in mind the desire for generality, the structure should be rich enough in detail to be a good representation of not only other industrial economies but also the underdeveloped and developing countries. Fitting the model to a particular country would merely require selection of suitable parameters and initial conditions. However, all present work is in terms of the United States.

A. Overview

The model will treat all major aspects of the socio-economic system as internal variables to be generated by the interplay of mutual influences within the model structure. The model will contain production sectors, labor and professional mobility between sectors, a demographic sector with births and deaths and with subdivision into age categories, commercial banking to make short-term loans and generate credit, savings institutions to accept saving and to make long-term loans, a monetary authority with its controls over money and credit, government services, government fiscal operations, consumption sectors, and a foreign sector for trade and international monetary flows.

A generalized production sector is being created with a structure comprehensive enough that it can be used, with selection of suitable parameters, for each of some fourteen or more producing sectors in the economy. Each sector will reach down in detail to some eleven factors of production, ordering, and inventories for each factor of production, marginal productivities for each factor, balance sheet and profit and loss statement, output inventories, delivery delay computation, production-rate planning, price setting expectations, and borrowing.

The model is being formulated for the new DYNAMO III compiler, which handles arrays of equations and makes especially easy the replication of the production sector and its subparts. For example, an equation in the ordering function need be written only once with array subscripts to identify the ordering functions for each factor and sector.

When fully developed, the model will contain some 2000 level variables (referred to variously as integrations, state variables, stocks, or accumulations). This compares with 22 level variables in the Urban Dynamics model and 5 in World Dynamics. The total number of defined variables are about six times the number of level variables.

By reaching from national monetary and fiscal policy down to ordering and accounting details within an individual production sector, the model will bridge between the concepts of macro-structure and micro-structure in the economic system. We believe that the major modes of the economy arise from such a depth of structure and that highly realistic and informative behavior should emerge from such a degree of disaggregation.

B. Standard Production Sector

A standard production sector will be replicated to form a major part of the model. By choosing suitable parameter values, the standard sector can be used for consumer durable goods, consumer soft goods, capital equipment, building construction, agriculture, resources, energy, services, transportation, secondary manufacturing, knowledge generation, self-provided family services, military operations, and government service. Such generality focuses attention on the fundamental nature of production of goods and services and simplifies both construction and explanation of the model.

Within each production sector are inventories of some eleven factors of production--capital, labor, professionals, knowledge, energy, services, buildings, land, transportation, and two kinds of materials. In addition, production is affected by length of work week for labor, length of work week for capital, and the content of each of the two kinds of material in the product.

For each factor of production, an ordering function will create an order backlog for the factor in response to desired production rate, desired factor intensity, marginal productivity of the factor, price of the factor, price of the product, growth expectations, product inventory and backlog, profitability, interest rate, financial pressures, and delivery delay of the factor. In terms of dynamic behavior, the ordering function will be far more influential than the production function, yet, in the economics literature, attention has been in the reverse priority.

The structure of a standard production sector is essentially the structure of a single firm in the economy with parameters and nonlinear relationships chosen to reflect the broader distributions of responses resulting from aggregating together the many firms within a sector. As with a firm, the sector will have an accounting section that pays for each factor of production, generates accounts receivable and payable, maintains balance sheet variables, computes profitability, saves, and borrows money. The structure should generate the full range of behavior that arises from interactions between the real and the money and information variables. By carrying the model to such detail, it should communicate directly with the real system where a wealth of information is available for establishing the needed parameter values.

A sector will generate product price in accordance with conditions within the sector and between the sector and its customers. For testing price and wage controls, coefficients are available to inhibit price changes. The sector will distribute output among its customer sectors. Market clearing, or the balance between supply and demand, will be struck not by price alone but also on the basis of delivery delay reflecting availability, rationing, and allocation.

C. Labor and Professional Mobility

People in the production sectors are divided into two categories--labor and professional. For each category a mobility network defines the channels of movement between sectors in response to differentials in wages, abailability, and need. A mobility network has a star shape with each point ending at a production sector and terminated in the level representing the number of people working in the sector. At the center of the star is a general unemployment pool, which is the central communication node between sectors. Between the central pool and each sector is a "captive" unemployment level of those people who are unemployed but who still consider themselves a part of the sector. They are the people searching for better work within their sector or who are on temporary layoff but expecting to be rehired. In a rising demand for more labor, those in the captive level can be rehired quickly but longer time constants are associated with drawing people from other sectors by way of the general unemployment pool.

D. Demographic Sector

The demographic sector generates population in the model by controlling the flows of births, deaths, immigration, and aging. Age categories divide people into their different roles in the economy from childhood through retirement. The demographic sector divides people between the labor and professional streams in response to wages, salaries, demands of the productive sectors, capacity of the educational system, and family background. Workforce participation determines the fraction of the population working in response to historical tradition, demand for labor, and standard of living.

E. Household Sectors

The household sectors are replicated by economic category--labor, professional, unemployed, retired, and welfare. Each household sector receives incomes, saves, borrows, purchases a variety of goods and services, and holds assets. Consumption demands respond to price, availability of inputs, and the marginal utilities of various goods and services at different levels of income.

F. Financial Sector

The financial sector is divided into three parts--commercial banking, savings institutions, and the monetary authority. The financial sector determines interest rates on savings and bonds, buys and sells bonds, makes long-term and short-term loans, and creates intangible variables like confidence in the banking system.

The commercial banking system receives deposits, buys and sells bonds, extends loans to households and businesses, and generates short-term interest rates. In doing so it manages reserves in response to demands of the monetary authority, and acts in response to discount rate, expected return on investment portfolio, demand for loans, and liquidity needs.

The savings institution receives savings, extends long-term loans to households and businesses, generates long-term interest rates, buys and sells bonds, and borrows short-term from the banking system. The savings institution balances money, bonds, deposits, and loans. It allocates loans between businesses and households, and it monitors the debt levels and borrowing capability of each business and household sector.

The monetary authority controls discount rate, open market bond transactions, and required reserve ratios. In doing so it responds to such variables as owned and borrowed reserves of the bank, demand deposits, inflation rate, unemployment, and interest rates.

VI. STATUS, SCHEDULE, PROCEDURE

Phase One of the project has been almost entirely devoted to completing a preliminary formulation of the model that can become the basis for discussion and improvement. Such a preliminary model can be a powerful communication medium for eliciting inputs from experts. Because the model clearly reveals the assumptions made in the preliminary formulation, those with different or additional perceptions of the actual system can immediately identify errors and omissions. Preliminary model formulation is now nearly complete. At the present time (March 1975) assembly is under way and should be complete by August.

Phase Two from September 1974 through August 1975 includes a search for advisors and participants from whom to solicit suggestions for reformulation and improvements. The structure of the model and the nature of the DYNAMO III compiler permit very easy modification of the model. Because the model is so much better for communication purposes after a preliminary set of equations are operating, the most efficient time to use outside advice, criticism, and suggestions is after the preliminary model is usable. Individuals and organizations are now being identified with whom to interact to improve the model and to begin interpreting its implications.

Phase Three will extend through the last third of 1975 and possibly for a decade beyond. During Phase Three a widening circle of participants should become involved in a progression of discussions, model modifications, and publications on structure, behavior, and implications. As sufficient confidence in the model develops, the national issues to which it is addressed will be explored.

REFERENCES

1. Forrester, Jay W. Industrial Dynamics, MIT Press, Cambridge, Mass., 1961.

2. Forrester, Jay W. Urban Dynamics, MIT Press, Cambridge, Mass., 1969.

3. Forrester, Jay W. World Dynamics, Wright-Allen Press, Inc., Cambridge, Mass., 1971.

4. Oltmans, Willem L., On Growth, G. P. Putnam's Sons, New York, 1974.

5. Forrester, Jay W., Gilbert W. Low, and Nathaniel J. Mass. "The Debate on World Dynamics: A Response to Nordhaus," Policy Sciences, vol. 5, June, 1974.

6. Forrester, Jay W. "Educational Implications of Responses to System Dynamics Models," Management Science, forthcoming.

7. Forrester, Nathan B. The Life Cycle of Economic Development, Wright-Allen Press, Inc., Cambridge, Mass., 1973.

8. Forrester, Jay W. Principles of Systems, Wright-Allen Press, Inc., Cambridge, Mass., 1971.

9. Senge, Peter M. "An Evaluation of Generalized Least Squares Estimation," System Dynamics Working Paper D-1944-6, December 6, 1974, Alfred P. Sloan School of Management, Massachusetts Institute of Technology.

10. Mass, Nathaniel J. Economic Cycles: An Analysis of Underlying Causes, Wright-Allen Press, Inc., Cambridge, Mass., 1975.

THE DYNAMICS OF ECONOMIC FLUCTUATIONS: A FRAMEWORK FOR ANALYSIS AND POLICY DESIGN

Nathaniel J. Mass
Assistant Professor of Management
Alfred P. Sloan School of Management
Massachusetts Institute of Technology, Cambridge, Massachusetts

I. INTRODUCTION AND SUMMARY

This paper describes a general framework for assessing the validity of alternative theories of business cycles and longer-term economic cycles.* The evaluative framework draws upon a generic model of the production sector of the economy. As discussed in Section III of this paper, the production sector model (termed the "basic production sector") differs intrinsically from conventional theories of the firm by incorporating conserved levels of inventories and backlogs which decouple rates of ordering, shipments, and production; inclusion of such "buffer stocks" permits treatment of disequilibrium behavior characteristic of economic cycles. Evaluation of alternative theories of economic cycles is carried out by isolating the central causal elements of each theory and incorporating those elements into the basic production sector. Through computer simulation of the resulting model, the relative importance of each hypothesized factor can be gauged. The results of such evaluation should have great value to economic theorists and policy-makers who are currently confronted with a morass of partially overlapping and partially conflicting theories of cyclic economic behavior. Clarification of the underlying causes of economic cycles, in turn, should contribute to the design of enhanced policies for economic stabilization.

Section II of this paper expands upon the motivation for evaluating extant theories of economic cycles. Section III outlines the principal components of the production sector model used in the evaluation. Section IV applies the evaluative framework to study the role of labor adjustments and fixed capital investment in generating economic cycles. As noted in Section IV, the majority of business-cycle theories, including, for example, those of John Hicks, Paul Samuelson, and James Duesenberry, center on the role of fixed capital investment in the short-term business cycle. However, Section IV demonstrates that the long delays inherent in planning, construction, and depreciation of capital equipment render it unlikely that investment in fixed capital is an intrinsic cause of short-term cycles. Instead, the analysis suggests that labor-acquisition and short-term production and inventory-management policies principally underlie typical four-year business cycles. Moreover, capital-investment policies appear to cause an eighteen- to twenty-year cycle in potential output resembling the so-called "Kuznets cycle." Finally, Section V summarizes the paper and outlines directions for future theoretical and policy-oriented research.

II. THE NEED FOR CRITICAL EVALUATION OF BUSINESS-CYCLE THEORIES

Empirical studies of the national economy have identified several cyclical fluctuations characteristic of aggregate economic behavior. Of these, the "business cycle" is a recurring fluctuation of approximately four years' duration in output, prices, investment, and unemployment. In the intermediate range of periodicities, the "Kuznets cycle" is an eighteen- to twenty-year fluctuation in the rate of growth of capital stock and, therefore, in "potential output."** Finally, at the long end of the spectrum, the "Kondratieff cycle" is a fifty-year cycle in prices, interest rates, and capital investment.***

The volume of empirical research and theoretical study that has been devoted to each of the three principal economic cycles described above varies widely. For example, beginning chiefly with the work of Burns and Mitchell (1946), a wide range of empirical evidence has been assimilated regarding the timing, periodicity, and phase relationships characteristic of short-term business cycles. Numerous theories of the short-term cycle have also been advanced. In contrast to the large number of business-cycle studies, Abramovitz (1961) has noted that relatively little data or theory describing the causes and properties of Kuznets cycles is currently available. Furthermore, evidence on the Kondratieff cycle is so sparse as to call into question the very existence of the hypothesized fifty-year cycle.†

The evaluative framework developed in this paper is designed to clarify the causes of the business cycle and the longer-term economic cycles. Clarification of the causes of business-cycle behavior is needed precisely because such a large number of business-cycle theories have been proposed. As Haberler has observed, "the analysis of existing theories of the cycle has furnished a number of hypotheses. Few of these seem to be definitely wrong or a priori impossible. What is unsatisfactory, however, is the exclusiveness with which many writers proclaim one or other of these hypotheses as the only possible solution."‡ One objective of the present work, therefore, is a first effort to integrate the existing theories of the business cycle in a unified framework of analysis.

A related objective is the exploration of the dynamic implications of widely held business-cycle theories. Are the economic decisions embodied in

---

*This research is supported under a grant provided by the Rockefeller Brothers Fund. The present paper is a distillation of the author's doctoral dissertation, "Generic Feedback Structures Underlying Economic Fluctuations," Alfred P. Sloan School of Management, Massachusetts Institute of Technology (December 1974). A revised version of the dissertation is forthcoming under the title, Economic Cycles: An Analysis of Underlying Causes (Cambridge: Wright-Allen Press, 1975).

**Potential output is defined as the maximum output derivable from the economy's existing stocks of labor, capital, and other factors of production. See Okun (1962) for further definition and discussion.

***For detailed discussion of the characteristics of the three cycles, see Gordon (1961), Hansen (1951), and Lee (1963).

†See, for example, Garvy (1943).

‡Haberler (1964), p. 361.

established business-cycle theories actually major determinants of the four-year cycle? Within the basic production sector, the relative importance of specified processes or relationships in generating short-term cycles can be tested. As discussed earlier, evaluation of alternative business-cycle theories in the proposed framework proceeds by incorporating the theories under examination into a common model structure. The relative validity and importance of each theory can then be assessed by analyzing the change in model behavior resulting when the structural elements under study are deleted or added.

To provide a concrete illustration of the theory-testing process, this paper utilizes the production sector model to study the role of labor adjustments and fixed capital investment in generating short-term business cycles. As noted by Burns (1969), the predominant number of business-cycle theories, including the theories of Paul Samuelson, John Hicks, Nicholas Kaldor, and James Duesenberry, emphasize fluctuations in fixed capital investment as a cause of overall fluctuations in income and output.* Such theories have been widely influential from a theoretical standpoint, and have stimulated much subsequent business-cycle research. In addition, widespread acceptance of the theories has led to formulation of economic stabilization policies designed to regulate investment opportunities. However, compared with other factors of production such as labor or in-process goods, fixed capital has a relatively long average lifetime, and is characterized by long lead times in planning, financing, and construction of new projects.** The long time constants associated with fixed capital investment and depreciation suggest that fixed capital variations cannot be a basic cause of short-term business cycles. Abramovitz has concisely presented a heuristic argument for this position:

> For a number of reasons, the simpler capital-stock adjustment models with their implied requirements for balanced growth rate take on heightened interest when considered in the context of long swings rather than in that of shorter business cycles. First, insofar as these models treat investment as dependent in part on current or past changes in the demand for finished goods, there has always been justifiable skepticism about their applicability to durable equipment and structures, so long as the theory was supposed to illuminate investment movements in short cycles. Since investment in durables is made for long periods of time it is doubtful whether it would respond readily to income change over short periods. This difficulty disappears, however, when we consider expansions lasting 8 to 12 years or more.***

According to Abramovitz, then, fixed capital investment is unlikely to be an essential factor in generating the short-term cycle, since the delays in capital formulation have about the same magnitude as the four-year business cycle and the delays in capital depreciation run much longer. The statement that fixed capital investment is not essential in generating the business cycle has two principal dimensions: first, that business cycles can occur independently of changes in fixed capital investment; and second, that fixed capital variations cannot independently generate four-year cycles. Section IV of the paper shows, for example, that, if fixed capital stock does not vary, short-term production and inventory-management policies governing labor acquisition still generate four-year cycles. Moreover, Section IV indicates that four-year business cycles do not appear in an economy where fixed capital is the only variable factor of production. These results demonstrate that fixed capital variations cannot be an intrinsic cause of four-year business cycles. More broadly, the results illustrate the use of the proposed evaluative framework for discriminating among alternative theories of the business cycle.

With regard to the longer-term economic cycles, analysis of the basic production sector within the proposed framework reveals that fixed capital investment may underlie the observed eighteen- to twenty-year Kuznets cycle.† Such a conclusion is consistent with the previous quotation from Abramovitz; moreover, the results broadly illustrate the potential of the proposed evaluative framework for generating insights about previously unstructured systems and unexplained modes of behavior.

III. OVERVIEW OF THE BASIC PRODUCTION SECTOR

Figure 1 provides a brief overview of the production sector model used in the evaluative framework.‡ The basic production sector is a generic model of a producing unit within a national economy. The sector receives orders which accumulate in an unfilled order backlog.

*See Samuelson (1939), Hicks (1949), Kaldor (1940), and Duesenberry (1958), for the original statement of these theories.

**Evans (1969) describes the delays in planning for and obtaining new fixed capital in terms of an "administrative lag" and an "appropriations lag." The administrative (decision) lag subsumes the time required to formulate actual investment plans; the appropriations lag intervenes between appropriation and actual investment expenditures. Evans estimates the sum of the two lags at about two years.

***Abramovitz (1961), p. 537, in R. A. Gordon and L. R. Klein, ed., Readings in Business Cycle Theory (1965).

†Mass (1975), Chapter 4, further suggests that capital-production behavior may underlie long-term cycles in the periodicity range of the Kondratieff cycle.

‡The production sector model is a simplified version of the production sector of the national socio-economic model under construction by the System Dynamics Group at MIT. For a description of the overall objectives and focus of the project, see Jay W. Forrester, "Understanding Social and Economic Change in the United States," Proceedings of the 1975 Summer Computer Simulation Conference, San Francisco (July 1975).

Based upon its production capacity and inventory and order backlog levels, the sector generates a shipment rate of output; the shipment rate depletes both inventory and order backlog. The production sector also formulates an output decision based on the adequacy of inventories and on the magnitude of the sector's order backlog. Desired production rate, in turn, along with desired factor proportions, governs the sector's desired inventories of factor inputs such as labor and capital.* The discrepancy between actual and desired factor inventories controls acquisition (ordering) of factor inputs. Finally, the sector's available factor inventories determine production rate.**

Figure 2 provides a more detailed view of the structure of the production sector. At the left of the figure, production rate is determined by the available stocks of labor and capital.+ Production rate adds directly to inventory and also affects shipment rate, with shipments rising with increased production capacity.‡

The production rate decision forms the nucleus of the production sector model. In the sector model,

(1) DPROD = APROD + (DINV - INV + BL - DBL)/TCIB

where

DPROD - Desired production (output units/year)
APROD - Average production (output units/year)
DINV - Desired inventory (output units)
INV - Inventory (output units)
BL - Backlog (output units)
DBL - Desired backlog (output units)
TCIB - Time to correct inventories and backlogs (years)

Equation (1) dictates expansion of output--an increase in production above the recent average production rate--whenever the sector's desired inventory exceeds available inventory or when backlogs are deemed excessively

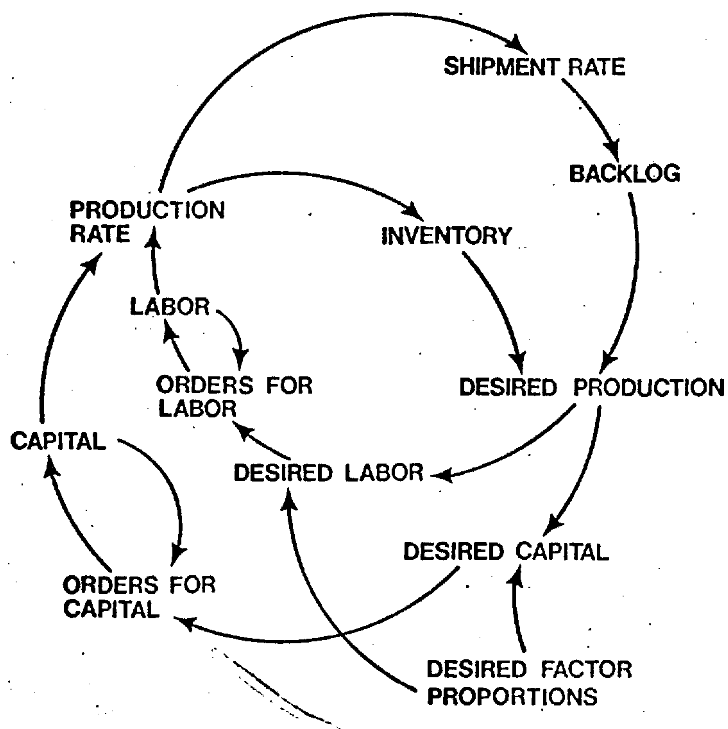


Figure 1. Feedback structure governing levels of labor and fixed capital.

*For simplicity, only two factors of production, labor and fixed capital, are considered in the analysis. Labor and capital are chosen as contrasting factors because labor is typically acquired readily and has a comparatively short lifetime within the firm while, as discussed previously, capital is a durable asset characterized by long planning and construction delays.

**As described in Mass (1975), Chapters 3 and 4, utilization of factor inputs is considered variable, depending on the balance of production and desired production within the sector.

+The production function actually utilized is a Cobb-Douglas function, modified to account for variations in capacity utilization. See Mass (1975), Chapter 3.

‡Shipment rate in the model is also affected by inventory and backlog levels. Therefore, for example, low inventories tend to curtail shipments (or, equivalently, raise delivery delays) while large order backlogs create pressures to expand shipments.

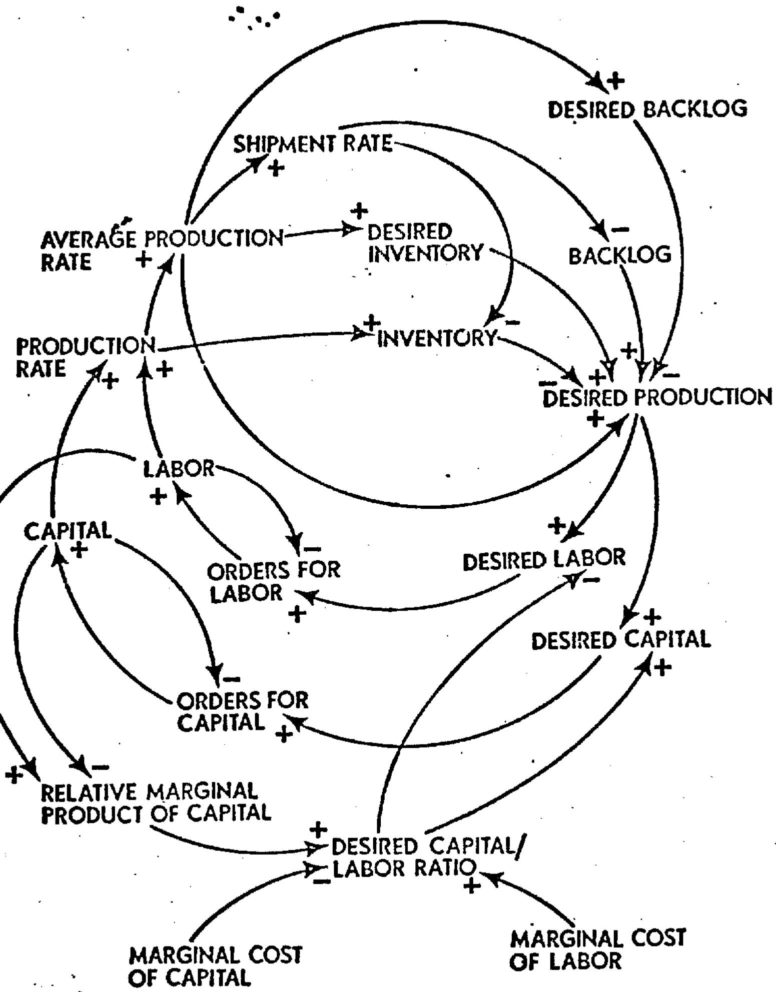


Figure 2. Simplified causal loop diagram of two-factor input model.

large. The coefficient TCIB controls the rate at which the sector attempts to eliminate inventory or backlog discrepancies.*

In the production sector, desired inventory DINV is assumed to equal average production rate APR multiplied by normal inventory coverage NIC. In the economics literature, desired inventories are typically assumed to depend on production or sales as a reflection of the level of activity of the firm. The motives for holding inventory have been described as follows by Ruth Mack:**

1. Bridging the time required for processes (economic transformations) to be performed
2. Efficient production or purchasing lots
3. Insurance against losing sales because of fluctuations in demand or other matters
4. Smoothing operations by provision for more or less foreseeable fluctuations
5. Grasping the potential advantage (or avoiding the disadvantage) of actual or expected changes in conditions in markets in which purchases or sales are made
6. Providing elective freedom from the tyranny of planning for uncertain events

Most of the six motives for holding inventory cited by Mack do not differentiate clearly the desirability of

---

*The inclusion of inventory and backlog correction terms in the equation for desired production appears consistent with results of recent econometric research. For example, Mack (1967) and Fromm (1961) conclude that desired inventories and unfilled orders both exert a significant impact on inventory investment. In addition, Darling (1959) found that inventory investment varies positively with sales and with changes in unfilled orders. Also, see Stanback (1962), p. 41, and Zarnowitz (1961), pp. 426, 451.

**Mack (1967), p. 27.

basing desired inventory on production rate versus sales rate. Certain of the motives, however, appear to favor relating desired inventory to production rate. For example, desired stocks of purchased materials and goods in process should depend on average production.* In addition, inventory ordering in most corporations is conducted by those directly involved in production. Therefore, in the production-sector equations, desired inventory is based on average production rate. This formulation may be regarded as an approximation to a more complex underlying structure that depends on both average production and sales.

Desired backlog DBL is assumed to equal average production rate APR multiplied by normal backlog coverage NBC. DBL therefore represents the order backlog that would prevail if production and shipment rate were both equal to average production and delivery delay were equal to a normal value of NBC years. The correspondence between normal backlog coverage and normal delivery delay can be seen in the definition of delivery delay:

Delivery delay = Backlog/Shipment rate.

The backlog correction term in Equation (1) therefore indicates whether, given the sector's current production capacity, orders in backlog can be filled in more or less time than the normal delivery delay. The sector attempts to expand capacity when backlogs are excessively large, and to contract capacity when backlogs are low.

The dynamic and behavioral significance of the inventory and backlog correction terms in Equation (1) for desired production merit discussion at this point. Dynamic models in the economics literature typically ignore the accumulation of production and orders in inventories and backlogs, respectively.** For example, rather than measuring the supply of goods by available inventory stocks, such models treat supply as a production flow rate governed by marginal costs of production. Analogously, demand is considered as a flow of purchases (consumption), rather than as a level of unfilled orders. Finally, price changes are typically assumed to occur whenever supply and demand, measured as rates of production and consumption, are imbalanced. Such a representation of supply and demand, however, fails to capture the disequilibrium behavior, and, in particular, the amplification, of production activity, characteristic of business cycles.†

To understand this deficiency of traditional economic models, consider the response of inventories and production within a firm or industry to a step increase in incoming orders.‡ As shown in Figure 3, as orders increase, pressures arise to expand production. The

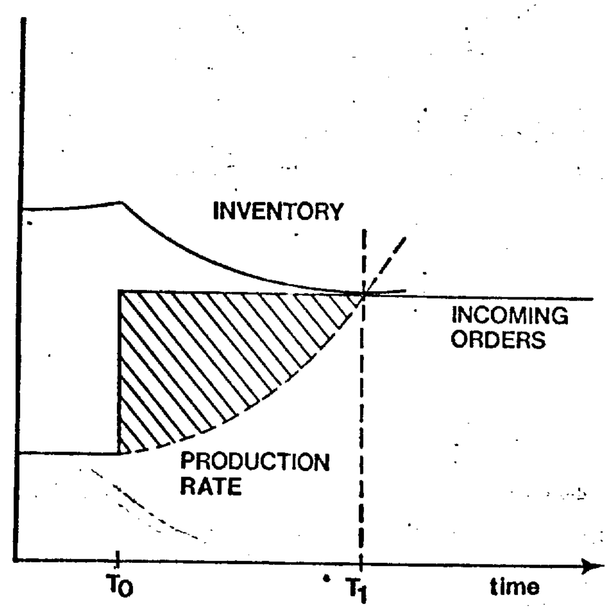


Figure 3. Inventory loss caused by increased sales.

*For example, Lovell (1961) bases desired stocks of purchased materials and goods in process on production rate in an econometric model of inventory investment.

**For example, see Samuelson (1939) and Arrow and Nerlove (1958).

†The term "amplification" refers, in the case of production, to the observed tendency for output to fluctuate by much more than sales or incoming orders over the typical business cycle. See Forrester (1961) for more precise definition. For empirical evidence on production behavior over the business cycle, see Gordon (1961), Haberler (1964), and Hansen (1951).

‡Note that the analysis in Figure 3 ignores backlog behavior for simplicity and, further, assumes that shipments (sales) equal incoming orders. Section IV elaborates upon this example.

increase in production rate is consistent with the classical economic analysis. However, consider point t_1 in Figure 3 where production has risen to equal consumption (orders). According to the classical model, supply and demand are equal at this point, and no further pressure on output or price should be exerted. However, all during the period t_0 to t_1 (in which production remained below incoming orders), inventories have been depleted. The total inventory loss over this interval is equal to the integral

$$\int_{t_0}^{t_1} (\text{Orders} - \text{Production})\, dt,$$

which measures the accumulated excess of consumption over production. In Figure 3, this integral is represented by the shaded area between the production and consumption curves. Therefore, at point t_1, supply equals demand in the rate-of-flow sense--production equals consumption--but inventory is below desired inventory.* Such an inventory imbalance would necessitate additional expansion of output beyond the incoming order rate, and would probably exert continued upward pressure on price. Therefore, the classical model of supply and demand which ignores inventory and backlog behavior fails to represent adequately the determinants of pricing behavior and, further, fails to capture the amplification and necessary overshoot of production engendered by changes in orders or consumption. These issues are discussed further in Section IV of the paper.

Referring once again to Figure 2, desired production, together with the desired capital/labor ratio, determines the desired stocks of capital and labor. The formulation resembles the neoclassical investment function described by Jorgensen (1967).** Comparison of the actual and desired stocks of capital and labor yields a discrepancy which modulates orders for the two factors. Finally, orders for factor inputs, after a delivery (acquisition) delay, add to the levels of capital and labor, respectively, thereby altering production capacity.***

IV. EXAMINATION OF THE ROLE OF LABOR ADJUSTMENTS AND CAPITAL INVESTMENT IN THE BUSINESS CYCLE

Section IV applies the evaluative framework outlined in Sections I-III to study the role of fixed capital investment and labor-hiring policies in generating short-term business cycles. Such an evaluation appears necessary because of the dominant emphasis in the literature and in policy debates on capital-investment theories of the short-term cycle. Section II developed the argument, however, that capital investment is unlikely to be an essential cause of the business cycle due to the relatively long delays associated with acquisition and depreciation of capital plant and equipment. As an alternative to the capital-investment theory, then, we might propose that variations in readily-acquired and relatively short-lived factors such as labor principally underlie the business cycle. In terms of the argument of Section II, the relatively short delays involved in labor recruitment and turnover suggest that labor adjustment could be an intrinsic cause of the cycle.

In order to investigate the above hypotheses, Section IV develops a sequence of three computer simulations. For the first simulation, labor is considered as the only variable factor of production. The resulting simulation exhibits a four-year cycle in production, employment (labor), and inventory. The second simulation, which considers capital as the only variable factor input, exhibits approximately a fifteen-year cycle in production rate and in the level of capital equipment. Finally, a third simulation, including both labor and capital as variable factors of production, exhibits the four-year production, employment, and inventory cycle characteristic of labor adjustments superimposed on a longer-term cycle in capital and potential output. This analysis verifies that the periodicities associated with adjustments in labor and fixed capital are sufficiently far apart that the individual cycles remain distinct when labor and capital combine as joint factors of production.

A. Economic Cycles Induced by Labor Adjustments

Figure 4 illustrates the response of the production sector, with labor as the only variable factor of production, to a 15% step increase in consumption (orders). This simulation is intended to isolate the periodicities associated with labor-hiring and termination policies. The consumption increase begins after one-half year (at time = .5).† Until time = .5, the system remains in equilibrium with production equal to the exogenous consumption (incoming order) rate of 3 million units per year. Also, labor, inventory, and backlog equal their desired values. Figure 4 exhibits approximately four-year fluctuations in labor, inventories, and production. The four-year period corresponds quite closely to the average 49-month period cited by Arthur Burns as characteristic of American business cycles.‡

In Figure 4, as consumption jumps from 3 million to 3.45 million units/year, backlog starts to rise; the

---

*Desired inventory would probably rise over the interval t_0 to t_1, reflecting both increased production and sales. The increase in desired inventory would tend to accentuate further the inventory discrepancy arising at time t_1.

**Also, see Jorgensen, Hunter, and Nadiri (1971) and Bischoff (1971). Mass (1975), Chapters 4 and 5, describes several differences between the "ordering function" utilized in the basic production sector and the neoclassical model.

***The delivery delay for factor inputs is not shown explicitly in Figure 2.

†As discussed by Forrester (1961) a simple test input such as a step input or random noise can excite the periodicities inherent in a system. Section 3.7 of Mass (1975) describes a model with endogenous consumption based on wage incomes. This section illustrates that multiplier-accelerator interactions discussed widely in the literature appear less fundamental in generating economic cycles than amplification produced by inventory- and backlog-management policies. See Samuelson (1939) for an overview of the multiplier-accelerator theory of business cycles.

‡The production sector model with labor as a single variable factor input exhibits a cycle that is consistently in the three- to five-year range when model parameters are varied by ±100%.

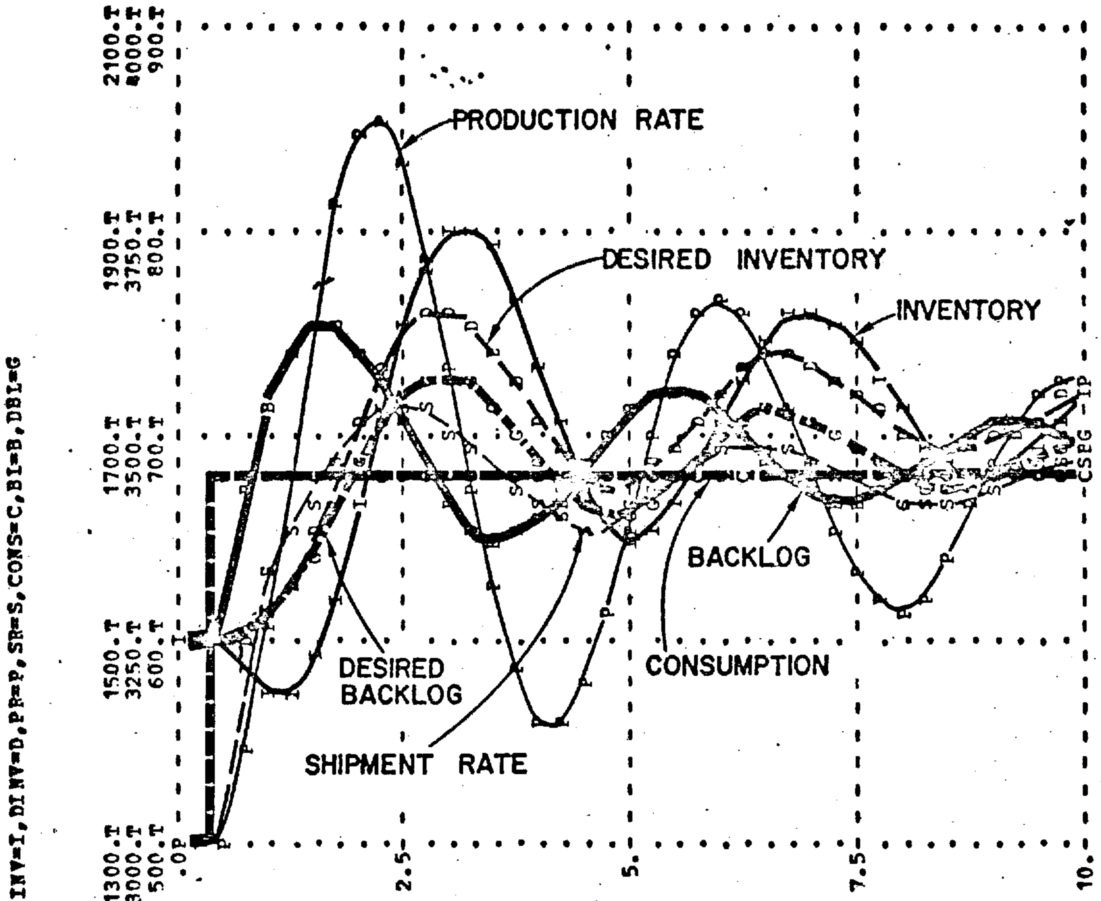


Figure 4(a)

Figure 4. Step response of basic production sector.

increased backlog raises the shipment rate, thereby depleting inventory. Inventory declines from an initial value of 1.5 million units to approximately 1.45 million units at the end of the first year. Over the same time, desired inventory rises to 1.55 million units. Backlog increases from .6 million units to about .75 million units at time 1.5, while desired backlog rises over the same interval from .6 million to .65 million units. The resulting divergence between actual and desired inventory and between actual and desired backlog causes a rapid expansion in desired production. Figure 4(b) shows desired production rising from 3 million units/year to a peak value of 3.6 million units/year at year 2. The rise in desired production causes a corresponding rise in desired labor from 1500 men to approximately 1850 men at year 2.

In Figure 4(a), production rises to a maximum value of about 3.9 million units/year in response to the increase in desired production. The increase in production at this point over the initial value of 3 million units/year is .9 million units/year, or double the increase in incoming orders. The peak in production slightly lags desired production due to the delays in acquiring labor. This lag is somewhat mitigated, however, by the use of overtime, instead of additional hiring, to expand production. For example, Figure 4(b) shows that the relative length of work week rises to a value of 1.07, implying a 7% increase in the average work week, at the start of the second year. Production rate exhibits a four-year period between years 2 and 6. The production cycle is caused by inventory and backlog policies that induce successive overshoot and undershoot of production relative to consumption.

In Figure 4(a), as long as the rate of production is below consumption and shipment rate, inventory continues to decline and backlog continues to accumulate. Consequently, inventory reaches a minimum value, and backlog approximately attains a maximum value at the point where production rises to just equal consumption.* This characteristic behavior of production rate

*In actuality, Figure 4(a) shows that backlog peaks roughly one-half year after production equals consumption. The slight lag occurs because shipment rate is constrained by low inventory; therefore, the shipment rate only equals consumption at time 1.5.

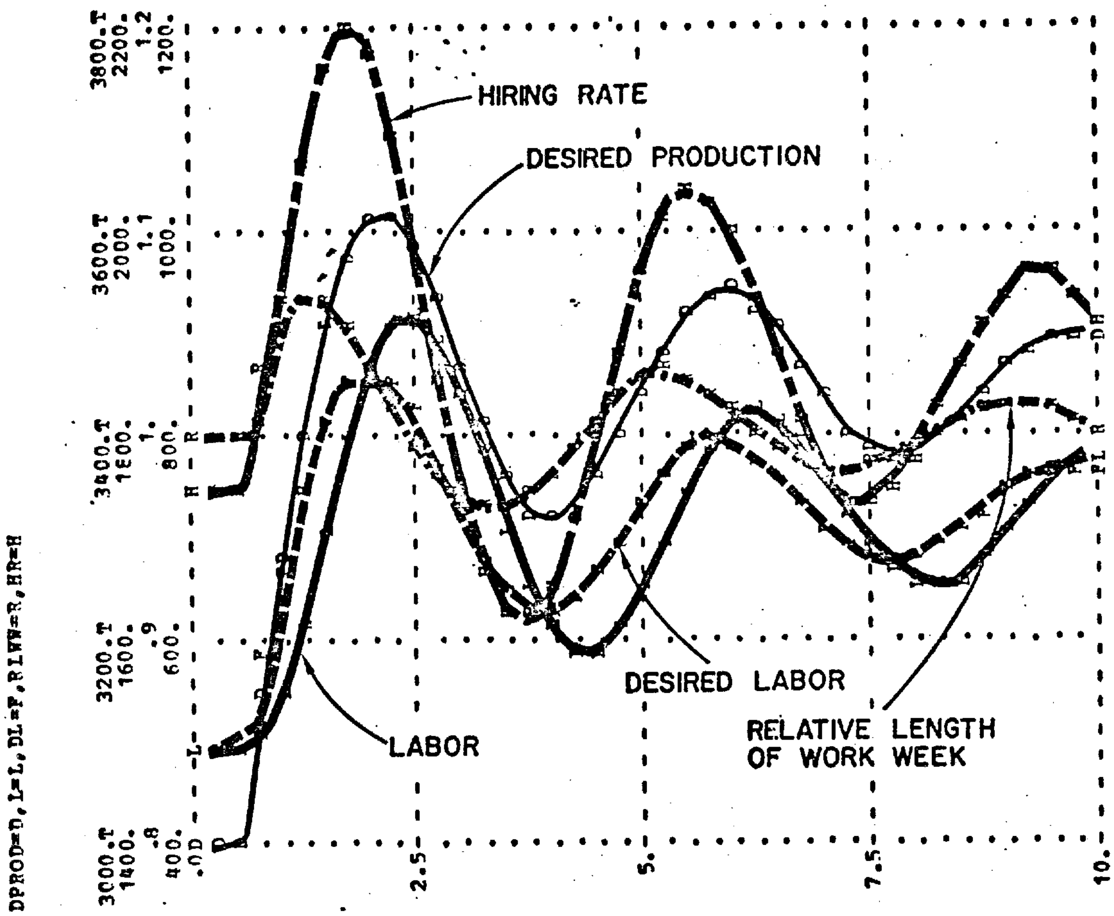


Figure 4(b)

and inventory is displayed in Figure 5. Figure 5 shows an expansion of production resulting from an increase in incoming orders occurring at time t_0; production rises until it equals the incoming order rate at time t_1. However, at time t_1, inventory is below, and order backlog above, their desired values. These discrepancies necessitate a continued expansion of production above the incoming order rate. In other words, even when production equals incoming orders, expansion must continue in order to eliminate the inventory shortage and large backlog accumulated while production was still less than the incoming order rate. For this reason, the computer output in Figure 4 shows that desired production rises above consumption at the start of the second year; desired production continues to remain above orders until approximately year 3.*

In a firm or in an entire economy, if labor and production continually expand as long as inventories are short, for example, the pattern illustrated in Figure 6 appears. In Figure 6, production expands in response to an increase in incoming orders until inventory builds up to equal desired inventory.** However, at the point where inventory equals desired inventory, production exceeds incoming orders. Inventory therefore continues to rise above desired inventory. The resulting inventory surplus can only be eliminated if production falls below incoming orders for some period of time. In this way, inventory adjustments lead to production fluctuations around the incoming order rate. Backlogs similarly exert a destabilizing effect on production rate, thereby accentuating the effect of inventories. Output must rise above incoming orders to eliminate the large backlogs accumulated during the initial upsurge in demand. The response of production to backlog behavior could be analyzed in a manner parallel to the response shown in Figure 6 for the case of inventories.

*For a detailed verbal and graphical analysis of the causes of inventory fluctuations in a simple second-order inventory-workforce model, see Mass and Senge (1974).

**Note that the swings in production would be accentuated if desired inventory is assumed to rise with increasing production or sales.

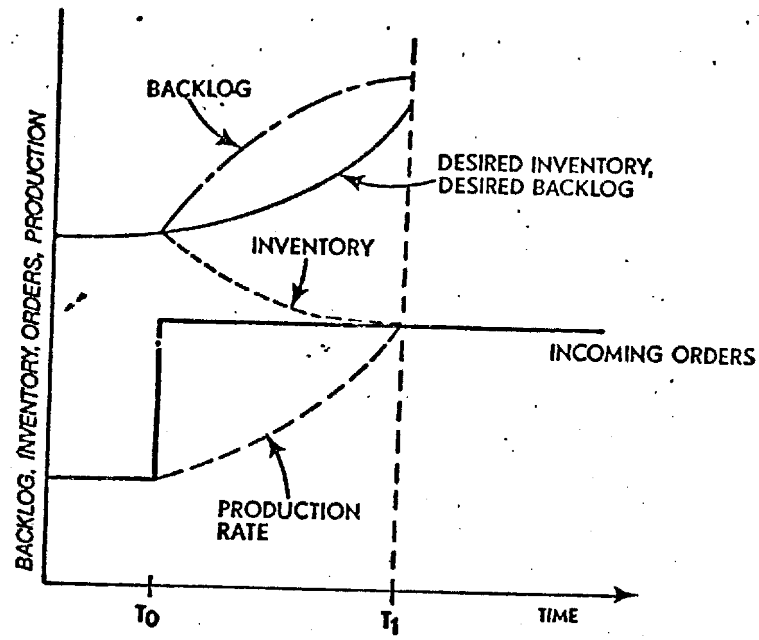


Figure 5. Inventory and backlog discrepancies arising from an increase in incoming orders.

The fluctuations in production in Figure 4 are convergent over time. Such convergent behavior is reasonable since the consumption rate is perfectly constant after the 15% step increase. The damping ratio characterizing production [defined as (1 - the ratio of successive peaks)] is approximately 55%, indicating that

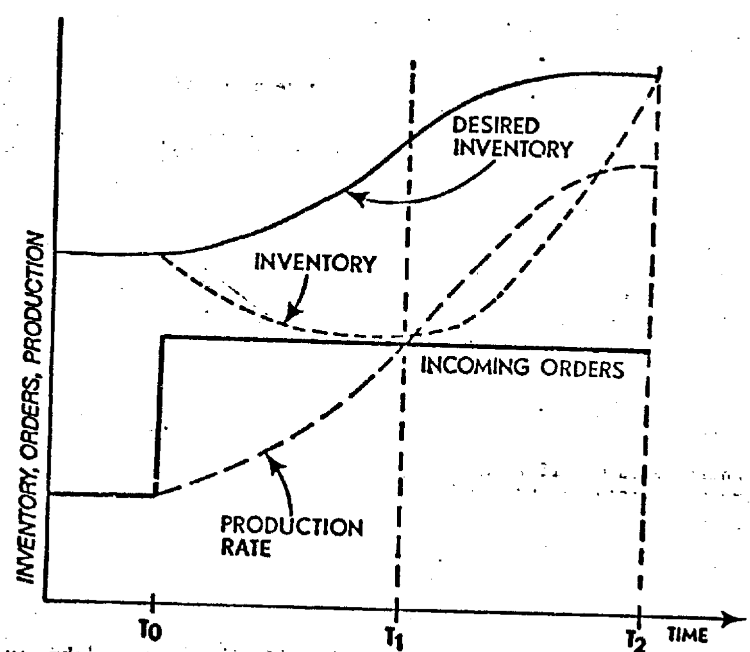


Figure 6. Increase of production above incoming orders.

55% of the overshoot of production is eliminated in each successive cycle.*

The phase relationships seen in Figure 4 appear to correspond closely with available statistical evidence on business cycles. For example, in Figure 4(a), backlog leads production by approximately one-half year. Backlog peaks before production since backlog begins to decline once shipment rate exceeds consumption; in contrast, production continues to expand beyond consumption, as discussed earlier, in order to build up inventory and reduce backlog to desired values. The one-half year lead of backlog with respect to production conforms closely to evidence presented in Zarnowitz (1961) and Stanback (1961).

Figure 4 also shows a slight lag--roughly one-quarter to one-half year in duration--of labor behind production. Production declines even while labor is increasing because use of overtime declines (seen in Figure 4(b) as a declining relative length of work week) as production begins to exceed desired production. Therefore, in Figure 4, hiring rate peaks at year 1.5, but labor continues to expand because the hiring rate still slightly exceeds the termination rate (not plotted). The brief lag of employment behind "reference cycle" peaks is discussed in Gordon (1961), p. 289.**

Finally, Figure 4(a) shows a one-year (quarter-cycle) lag of inventory behind production. Since shipment rate tends to lag production slightly, inventory lags shipment rate by about three-quarters of a year--a period very close to the 7-9 month lag cited by Abramovitz for manufacturing industries.*** Inventory tends to lag production because inventory continues to increase, even while production declines, as long as production exceeds shipment rate. Such a time-pattern of inventory behavior is observable in the aggregate economy as well as in many individual industries.

B. <u>Economic Cycles Induced by Fixed Capital Investment</u>

Figure 7 exhibits the response of the basic production sector including fixed capital as a single variable factor of production to a 15% step increase in consumption beginning at year 2.**** All system variables exhibit a cycle of approximately fifteen-year periodicity.***** Such a periodicity is well beyond the range of short-term business cycle fluctuations, but closely resembles the periodicities characteristic of long-term Kuznets cycles. The Kuznets cycle, according to Hickman, is a fifteen- to twenty-year fluctuation in the rate of growth of capital stock, output, productivity, and other variables.+ The Kuznets cycle is also characterized by long swings in growth of labor force and unemployment rate. Figure 8 illustrates average annual changes in output for the United States from 1860 to 1960. Figure 8 shows Kuznets-cycle peaks occurring roughly in 1865, 1885, 1900, 1920, and 1940.‡

Detailed analysis of the results in Figure 7 completely parallels the analysis of Figure 4, except that the response is drawn out over a much longer period and the magnitude of lead and lag relationships consequently differs widely. Nonetheless, a brief analysis is provided here to emphasize the parallels between causes of oscillations respectively induced by labor and fixed capital.

In Figure 7, as incoming orders increase, inventory begins to decline and order backlog increases. Desired production [plotted in Figure 7(b)] rises as a result, thereby leading to increased orders for capital. However, the lengthy planning and construction delays for fixed capital delay the increase in actual capacity acquisition. By around year 5, production has increased sufficiently to equal consumption, thereby terminating the drop-off in inventory. But inventory has been steadily depleted between years 1 and 5, while desired inventory has risen in response to increased production. At year 5, for example, inventory equals 1.25 million

---

*Production peaks at a value of 3.9 million units/year at time 2 and 3.65 million units/year at time 6. The damping ratio, defined relative to the mean point of the oscillation, is therefore:

$$1 - \frac{(3.65 - 3.45)}{(3.9 - 3.45)} \cong .55.$$

**Reference cycles have been analyzed by the National Bureau of Economic Research to identify turning points in general business activity. For a brief description of the approach, see Gordon, pp. 265-270.

***Abramovitz (1950), p. 119.

****Note that Figure 7 spans forty years on the horizontal scale.

*****The basic production sector including fixed capital also responds with a fifteen-year cycle to random noise in consumption. The result probably has greater practical significance than the step or ramp responses illustrated above, because random variation is necessarily superimposed on <u>all</u> consumption streams.

+Hickman (1963), pp. 490-492. The original empirical work supporting the existence of 18-20 year swings in capital growth appears in Wardwell (1927) and Kuznets (1930). Lewis and O'Leary (1955) have conducted a more recent study indicating the existence of Kuznets-type cycles in a variety of countries.

‡As mentioned above, Kuznets-cycle fluctuations appear in the rate of growth of output and capital stock. Such fluctuations superficially differ from the capital-production cycle illustrated in Figure 7 which exhibits a cycle in <u>absolute levels</u> of output and fixed capital. However, available data on the Kuznets cycle are drawn from a growing economy and therefore measure fluctuations in output and capital around a long-term growth trend. To produce comparable data, the basic production sector can be subjected to a steady ramp increase in incoming orders. Resulting simulations exhibit a fifteen-year cycle in capital growth rate, identical to the periodicity of capital in Figure 7.

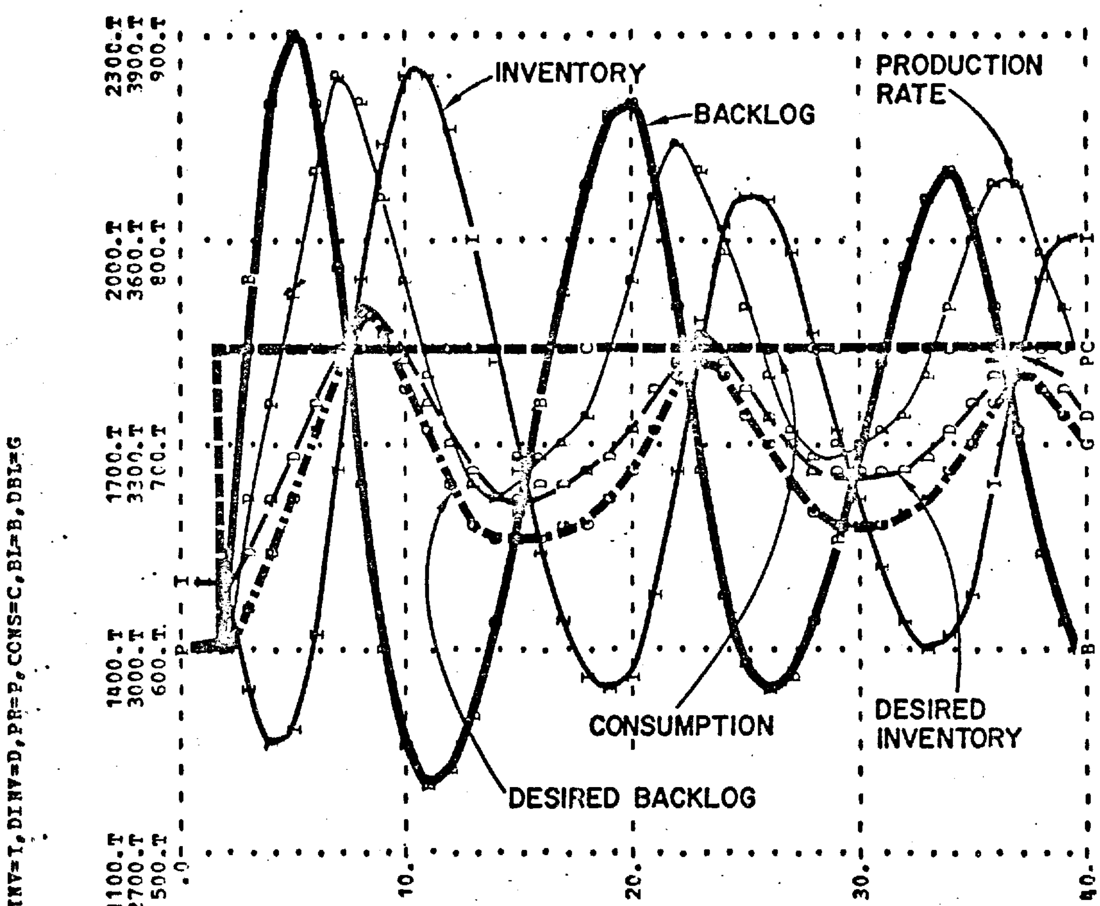


Figure 7(a)

Figure 7. Step response of basic production sector including fixed capital.

units while desired inventory equals nearly 1.7 million units [see Figure 7(a)]. Analogously, Figure 7(a) exhibits a large backlog discrepancy at year 5 with backlog equal to .9 million units and desired backlog equal to .66 million.

To eliminate the inventory and backlog discrepancies caused by increased consumption, production must rise above consumption. This, in Figure 7(b), capital rises from an initial value of 7.5 million units to about 10 million units at year 10. Capital continues to expand as long as desired production exceeds average production. However, as inventory builds up once more and backlog declines, desired production rate drops off, thereby gradually leading to excess capacity. Capacity remains in excess over several years as a consequence of the long delay in capital depreciation. Consequently, Figure 7 displays a cyclical adjustment similar to the behavior in Figure 4 where labor was considered the only variable factor of production. However, the response is protracted compared with the four-year cycles induced by labor adjustments.* Compared with Figure 4, addition of capacity is delayed on the production upturn due to increased acquisition delays; moreover, capacity is slowly reduced on the downturn as a result of gradual "runoff" of capital through depreciation.

The capital and production cycle in Figure 7, although considerably longer than the four-year business cycle, probably still lies on the short range of cycles induces by fixed capital investment. For example, the production sector underlying Figure 7 resembles an industrial sector producing consumer goods. A consumer-goods sector is characterized by a relatively short delivery delay for its output and, also, by a relatively short delay for in-process inventory between initiation and completion of production. Consequently, a consumer-goods sector normally experiences short delays in adjusting production to consumption. In order to study the behavior modes characteristic of a capital-producing sector, rather than a goods sector, the parameter values of the sector model can be adapted to describe a capital sector. Compared with the goods-producing sector, the capital sector is assumed to have a longer manufacturing and delivery delay for its output. Due to the increased

*Pigou and D. H. Robertson both attributed the period of business cycles to the "gestation delay" or construction period of capital equipment. The results in Figure 7 indicate the role of such construction delays, in addition to planning and depreciation delays, in generating capital cycles. However, contrary to Pigou and Robertson, cycles induced by fixed capital investment appear to have a much longer period than the four-year trade cycle.

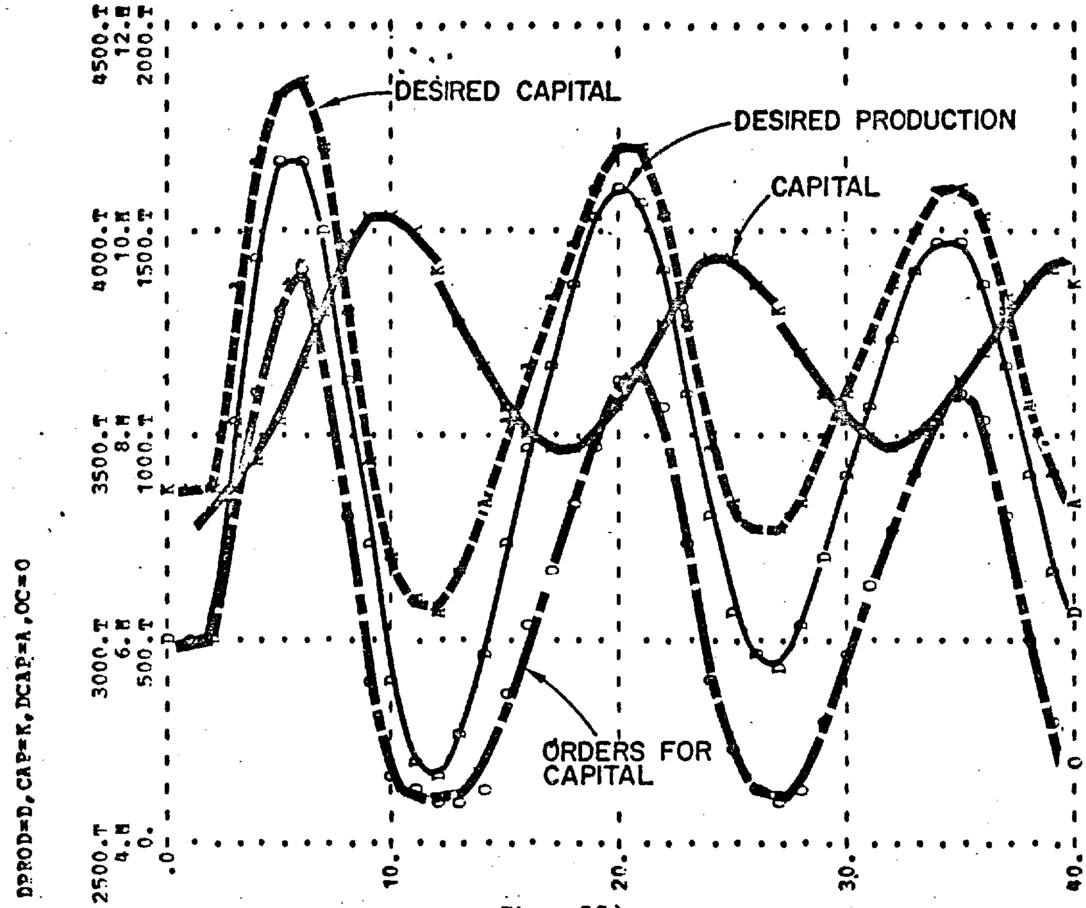


Figure 7(b)

delivery delay, the capital sector also takes longer to correct inventory and backlog discrepancies. Moreover, the capital sector is assumed to take longer to adjust production due to fears of overbuilding capacity if adjustment proceeds too rapidly; such risks of overexpansion tend to increase rapidly as the delivery delay of the sector increases. When the appropriate parameter changes are incorporated, the production sector exhibits roughly a twenty-year capital cycle.* In a model of the economy containing both goods-producing sectors and capital-producing sectors, the long-term capital cycles characteristic of the individual sectors would probably be mutually entrained to form a single capital cycle of around eighteen-year periodicity.**

In terms of the discussion of Section II, the results described here cast doubt upon the validity of any theory centered around fixed capital investment as an essential cause of the short-term businesss cycle. Instead, the results suggest that labor adjustments chiefly underlie short-term cycles in output and employment, while fixed capital investment generates longer-term cycles in growth of capital stock.

The practical and theoretical significance of these issues is well described by Gordon:

> Economists have not yet developed a generally accepted explanation of these intermediate swings, nor is there full agreement that these swings constitute a separate order of cycles distinct from business cycles. One uncertainty arises from the fact that these "cycles" are obviously related to the severe depressions of the past century. It is not surprising that expansion should be particularly

---

*Mass (1975), Chapter 4.

**The capital cycle exhibited by the aggregate economy might even be much longer than eighteen to twenty years in a model which contained structural elements missing from the basic production sector analyzed in Figures 4 and 7. For example, a model containing a limited supply of labor should tend to exhibit a longer-term capital cycle, since more of the burden of adjusting production to desired production would be accomplished through capital ordering rather than through short-term changes in employment. Incorporation of a limited labor supply in the sector model could well extend the period of the long-term capital cycle close to the fifty-year span of the Kondratieff cycle. Examination along these lines is currently being conducted within the framework of the basic production sector.

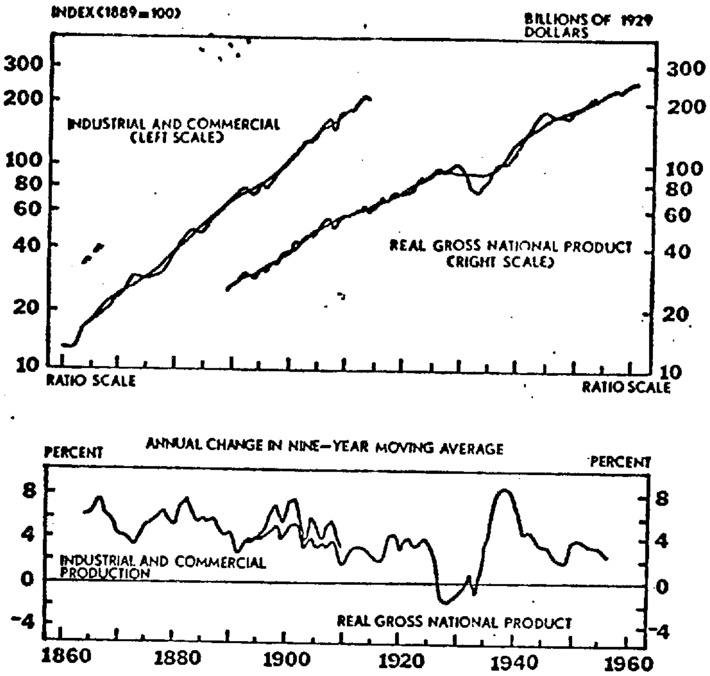


Figure 8. Long swings in aggregate production in the United States.

(Source: Hickman (1963), p. 491)

> rapid as the economy comes out of a deep depression, and the "downswings" of these long cycles may reflect in part the fact that we have experienced severe depressions. It is significant, however, that in the past, deep depressions have been associated with substantial retardation in the rate of growth of output.*

From a theoretical standpoint, the results presented above suggest a common structure underlying both business cycles and longer-term capital cycles. Moreover, according to this analysis, differences in characteristics of factors of production--differences such as average factor lifetimes and delivery delays which can be represented simply in terms of changed parameter values--suffice to explain the different periodicities of fluctuation.

C. <u>Economic Cycles Induced Jointly by Labor and Fixed Capital</u>

A final computer simulation, shown in Figure 9, integrates the preceding analysis. Previous simulations (Figures 4 and 7) dealt with capital and labor individually in order to study the cyclical modes arising from each factor input. The simulation in Figure 9 combines labor and capital in a joint production process. Figure 9 primarily investigates the question: "When capital and labor are combined, do the periodicities associated with each input factor remain distinct or are they mutually entrained to yield a single cycle of intermediate length?"

Figure 9, which plots fixed capital stock, labor, and production rate over a one-hundred-year period, clearly illustrates the different periodicities associated with labor and fixed capital; these results are in accordance with the analyses conducted previously on models containing single factors of production. The results lend further support to the hypothesis that labor adjustments principally underlie short-term business cycles and that fixed capital investment is not an intrinsic factor in generating business cycles.**

---

*Gordon (1961), p. 243.

**The business cycle is in fact characterized by short-term fluctuations in capital spending. Such observed fluctuations are consistent with the behavior of the revised production sector model (orders for capital were not plotted in Figure 9). According to the model, short-term investment cycles are the result of fluctuations in the relative balance of capital and desired capital within the economy. Fluctuations in capital relative to desired capital stock in turn reflect short-term changes in the balance of production and desired production caused by corporate policies governing overtime and labor adjustment. In other words, short-term employment and overtime policies adjust production rate towards desired production over the business cycle, thereby creating varying incentives for capital investment. However, the resulting short-term fluctuations in capital investment are directly caused by the labor policies; capital investment policies still cannot independently generate short-term business cycles. Future examination should aim at clarifying further the interactions of production, employment, and investment policies over the business cycle.

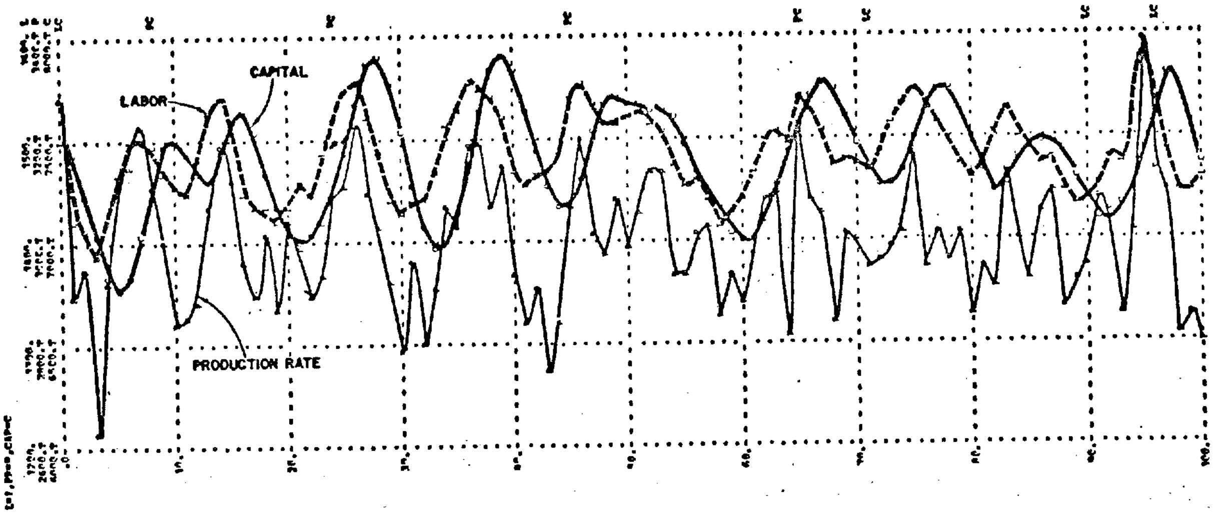


Figure 9. Response of two-factor model to noise in production rate--100 years.

D. Implications for Business-Cycle Theory and Stabilization Policy

The analysis conducted here generally indicates the importance of reassessing the fundamental assumptions underlying business-cycle theory and stabilization policy. Many prevalant economic stabilization policies, particularly monetary policies, are largely predicated on a capital-investment theory of business-cycle behavior. However, if business cycles are attributable for the most part to short-term employment and inventory decisions, policies that attempt to control fixed capital investment may have relatively little leverage or at least may be less effective than policies directly aimed at employment and inventories. Moreover, if fixed capital investment generates fifteen- to twenty-year or longer cycles in capital plant, policies designed to regulate capital investment can have significant long-term impacts on output, employment, and productivity.

The results presented here, although speculative, suggest the need for critical assessment of proposed economic stabilization policies according to (1) their short-term impacts on labor and inventory adjustments; and (2) their longer-term effects on capital investment and potential output. Such an evaluation may contribute to greater understanding of the impacts and probable effectiveness of discretionary monetary and fiscal policies and the various "automatic stabilizers."

V. CONCLUSIONS

This paper has developed a general framework for analyzing the validity of alternative theories of business cycles and longer-term economic cycles. The framework draws upon a general model of production activity interrelating inventories, backlogs, acquisition of factor inputs, and output. The production sector model has been used here to analyze the periodicities associated different economic factors of production. The results indicate, contrary to the prevalent capital-investment theories of the business cycle, that labor-adjustment policies, in conjunction with short-term production and inventory-management policies, appear to underlie the four-year business cycle; moreover, capital investment policies appear to be principally involved in economic cycles of much longer duration. These results motivate analysis of current and proposed economic stabilization policies in terms of their short-term effects on employment as well as their longer-term impacts on capital investment and potential output.

A wide range of economic processes can potentially be analyzed within the framework of the basic production sector. For example, extending the sector to include a simple monetary sector may help to clarify current debates over the role of money and interest rates in business cycles.* Detailed analysis along these lines should strengthen the foundations of business-cycle theory and, as a byproduct, enhance the exercise of stabilization policy.

*For a statement of the monetary theory of business cycles, see Wicksell (1907, 1935), Hayek (1935), and Friedman and Schwartz (1963).

VI. REFERENCES

1. Abramovitz, Moses. Inventories and Business Cycles (New York: National Bureau of Economic Research, 1950).

2. Abramovitz, Moses. "The Nature and Significance of Kuznets Cycles," in Robert A. Gordon and Lawrence R. Klein, eds., Readings in Business Cycle Theory (American Economics Association, 1965).

3. Arrow, Kenneth J., and Marc Nerlove. "A Note on Expectations and Stability," Econometrica, vol. 26 (April 1958), pp. 297-305.

4. Bischoff, Charles W. "Business Investment in the 1970s: A Comparison of Models," Brookings Papers on Economic Activity (1971:1), pp. 13-64.

5. Bronfenbrenner, Martin, ed. Is the Business Cycle Obsolete? (New York: John Wiley & Sons, 1969).

6. Burns, Arthur F., and Wesley C. Mitchell. Measuring Business Cycles (New York: National Bureau of Economic Research, 1946).

7. Burns, Arthur F. The Business Cycle in a Changing World (New York: National Bureau of Economic Research, 1969).

8. Darling, Paul G. "Manufacturer's Inventory Investment, 1947-1958," American Economic Review, vol. 49 (December 1959), pp. 950-962.

9. Duesenberry, James S. Business Cycles and Economic Growth (New York: McGraw-Hill, 1958).

10. Evans, Michael K. Macroeconomic Activity (New York: Harper & Row, 1969).

11. Forrester, Jay W. Industrial Dynamics (Cambridge: MIT Press, 1961).

12. Friedman, Milton, and Anna J. Schwartz. "Money and Business Cycles," Review of Economics and Statistics, vol. 45 (February 1963), pp. 32-78.

13. Fromm, Gary. "Inventories, Business Cycles, and Economic Stabilization," Part IV in Inventories and Economic Stabilization (Washington, D.C.: Joint Economic Committee, 1961), pp. 37-91.

14. Garvy, George. "Kondratieff's Theory of Long Cycles," Review of Economic Statistics, vol. 25 (November 1943), pp. 203-220.

15. Goodwin, R. M. "The Non-Linear Accelerator and the Persistence of Business Cycles," Econometrica, vol. 19 (January, 1951), pp. 1-17.

16. Gordon, Robert A. Business Fluctuations (New York: Harper & Row, 1961).

17. Gordon, Robert A., and Lawrence R. Klein, eds. Readings in Business Cycle Theory (Homewood, Ill.: Richard D. Irwin, Inc., 1965).

18. Haberler, Gottried. Prosperity and Depression (London: George Allen & Unwin, Ltd., 1964).

19. Hansen, Alvin H. Business Cycles and National Income (New York: W. W. Norton, 1951).

20. Hayek, F. A. Prices and Production (London: George Routledge, 1935).

21. Hickman, Bert G. "The Postwar Retardation: Another Long Swing in the Rate of Growth?", American Economic Review, vol. 53 (May 1963), pp. 490-507.

22. Hicks, John R. "Mr. Harrod's Dynamic Theory," Economica, vol. 16 (May 1949), pp. 106-121.

23. Hicks, John R. A Contribution to the Theory of the Trade Cycle (London: Clarendon Press, 1950).

24. Jorgensen, Dale W., and Calvin Siebert, "A Comparison of Alternative Theories of Corporate Investment Behavior," American Economic Review (September 1968), pp. 681-712.

25. Jorgensen, Jerald Hunter, and M. Ishag Nadiri. "A Comparison of Alternative Econometric Models of Quarterly Investment Behavior," Econometrica, vol. 38 (March 1970), pp. 187-212.

26. Kaldor, Nicholas. "A Model of the Trade Cycle," Economic Journal, vol. 50 (March 1940), pp. 78-92.

27. Keynes, John M. The General Theory of Employment, Interest and Money (New York: Harcourt, Brace, 1936).

28. Kondratieff, N. D. "The Long Waves in Economic Life," Review of Economic Statistics, vol. 17 (November 1935), pp. 105-115.

29. Kuznets, Simon. Secular Movements in Production and Prices (New York: Houghton Mifflin, 1930).

30. Lee, Maurice W. Macroeconomic Fluctuations, Growth and Stability (Homewood, Ill.: Richard D. Irwin, Inc., 1963).

31. Lewis, W. A., and P. J. O'Leary. "Secular Swings in Production and Trade, 1870-1913," The Manchester School of Economics and Social Studies, vol. 23 (May 1955), pp. 113-152.

32. Lovell, Michael C. "Factors Determining Manufacturing Inventory Investment," Part II of Inventory Fluctuations and Economic Stabilization (Washington, D.C.: Joint Economic Committee, 1961), pp. 119-158.

33. Mack, Ruth P. Information, Expectation, and Inventory Fluctuation (New York: National Bureau of Economic Research, 1967).

34. Mass, Nathaniel J. Economic Cycles: An Analysis of Underlying Causes (Cambridge: Wright-Allen Press, 1975).

35. Mass, Nathaniel J., and Peter M. Senge. Understanding Oscillations in Simple Systems," System Dynamics Group Memorandum D-2045-1 (July 1974).

36. Metzler, Lloyd A. "The Nature and Stability of Inventory Cycles," Review of Economic Statistics, vol. 23 (August 1941), pp. 113-129.

37. Okun, Arthur M. "Potential GNP: Its Measurement and Significance," 1962 Proceedings of the Business and Economic Statistics Section of the American Statistical Association.

38. Pigou, A. C. Industrial Fluctuations (London: Macmillan & Co., Ltds., 1927).

39. Robertson, D. H. A Study of Industrial Fluctuations (Westminster: P. S. King & Son, Ltd., 1915).

40. Samuelson, Paul A. "Interactions Between the Multiplier Analysis and the Principle of Acceleration," Review of Economics and Statistics, vol. 21 (May 1939), pp. 75-78.

41. Stanback, Thomas M., Jr. Postwar Cycles in Manufacturers' Inventories (Princeton, N.J.: Princeton University Press, 1961), portions reproduced in Inventory Fluctuations and Economic Stabilization (Washington, D.C.: Joint Economic Committee, 1961), Part II, pp. 1-143.

42. Wardwell, C. A. An Investigation of Economic Data for Major Cycles (Philadelphia: 1927).

43. Wicksell, Knut. "The Influence of the Rate of Interest on Prices," Economic Journal, vol. 17 (1907), pp. 213-220.

44. Wicksell, Knut. Lectures on Political Economy (London: Macmillan, 1935).

45. Zarnowitz, Victor. "The Timing of Manufacturers' Orders During Business Cycles," in Geoffrey H. Moore, ed., Business Cycle Indicators (Princeton, N.J.: Princeton University Press, 1961).

46. Zarnowitz, Victor, ed. The Business Cycle Today (New York: National Bureau of Economic Research, 1972).

WORKFORCE SKILL-COMPOSITION AND HIGHER EDUCATION IN THE NATIONAL SOCIO-ECONOMIC MODEL

Peter M. Senge
System Dynamics Group
Alfred P. Sloan School of Management
Massachusetts Institute of Technology, Cambridge, Massachusetts

I. BACKGROUND

This paper presents a simple dynamic model to investigate basic forces which alter the skill-composition of the national workforce. The model attempts to identify long-term causes for shifts out of "labor" skill-level jobs and into "professional" skill-level employment in an industrial country such as the United States. As the term is used here, a laborer is a worker who engages directly in the production of some good or in the provision of a service which does not require highly specialized training. Laborers therefore include skilled and unskilled craftsmen, operatives in manufacturing and services, and manual laborers in manufacturing and agriculture. Conversely, professionals include managers; salesmen; and proprietors; as well as doctors, lawyers, engineers, teachers, and persons in other types of employment who are commonly referred to as professionals.

The present model focuses on interactions between workforce composition and higher education. All persons who receive higher education (schooling beyond that regarded as normal during a particular time period) are regarded as professionals. In the United States, growth in the professional workforce has been accompanied by expansion of institutions of higher learning and increasing enrollments of young adults in such institutions. The workforce-composition model attempts to investigate feedback interactions between workforce composition and higher education. The model investigates the impacts of expansion in opportunities for higher education on the mix of laborers and professionals in the nation. The model also attempts to explain how, in turn, changing workforce composition generates increased or decreased demand for higher education.

The model focuses on long-term changes in two variables in particular. The "fraction of the workforce in professional" provides a measure of workforce composition. According to data presented by P. M. Blau and O. D. Duncan, the fraction of the workforce with professional skills in the United States rose from about .18 in 1900 to approximately .35 in 1960.* The model also attempts to generate plausible behavior for the fraction of the student-age workforce in higher education, called the "fraction in higher education." The US Office of Management and Budget estimates that the fraction of adults 18 to 24 years old enrolled in undergraduate education equaled .24 in 1970.** Although no historical data series on enrollments in higher education seems to exist, the fraction in higher education appears to have risen continuously over the past thirty years or more.

The model of workforce skill-composition summarizes certain issues being investigated in the current national modeling project being conducted in the System Dynamics Group at MIT. The national model attempts to explain short-, intermediate-, and long-term patterns of economic and social change in the United States over the period 1850 to 2100.+ The model presented in this paper isolates a small number of the long-term issues which are important within the scope of the national model.

The relative availability of labor and professional workers to producing sectors in the national model will affect the rate at which many model sectors can grow. Relative shortages of professionals may constrain growth in certain sectors during early stages of national growth, and shortages of laborers may be important during later stages of growth. Therefore, the long-term shift out of labor into professional may be a significant aspect in understanding growth of the national system.

The growth of the knowledge sector (representing colleges and universities, research institutes, and research and development divisions of firms) is also an important long-term dynamic in the national model. A major factor driving growth of the knowledge sector is demand for higher education. Therefore, the model presented below represents an initial attempt to understand interactions between demand for higher education and changing workforce composition, which must be treated in the national model. Understanding generated in the present model will contribute to how these interactions are represented in the national model.

The model presented below simplifies many complex interactions in the full national model. The model incorporates numerous exogenous variables, such as population, technological complexity, and relative demand for laborers and professionals in the nation, which are part of the feedback structure of the full national model. The model also assumes certain variables to be constant which in fact vary in the full national model. Such simplifying assumptions are made in order to permit the present model to focus on a small set of feedback interactions which appear to be important in altering workforce composition over the course of national development.

II. MODEL STRUCTURE

Section II presents a broad overview of the structure of the workforce composition model. The factors assumed to alter the mix of laborers and professionals in the workforce are described in Section II.A. Section II.A describes each factor altering workforce composition in the context of actual movements of workers in the economy. Although all factors affecting workforce composition described in Section II.A will ultimately be part of the feedback structure of the full national model, the present model focuses on feedback interactions between demand for higher education and workforce skill-composition. Section II.A also identifies factors altering the laborer-professional mix, such as promotions and immigrations, which are assumed to be exogenously

---

*Data estimated from P. M. Blau and O. D. Duncan, The American Occupational Structure, Wiley and Sons, 1967 (Figure 3.1, pp. 86-88).

**US Office of Management and Budget, Social Indicators, US Department of Commerce, 1973 (Figure 3/13, p. 87).

+See J. W. Forrester, "Understanding Social and Economic Change in the United States," Proceedings of the Summer Computer Simulation Conference, San Francisco, 1975.

determined in the present model. Section II.B presents the feedback structure which describes how the demand for higher education responds to changes in workforce skill-composition and how the fraction in higher education subsequently alters workforce composition.

II.A. Factors Altering Workforce Skill-Composition

The workforce composition model incorporates five distinct factors which cause changes in the mix of laborers and professionals in the nation. As shown in Figure 1, workforce composition changes in response to promotions of laborers to professional, movements of unemployed professionals back to labor jobs, immigrations, movements of (previously unpaid) farm workers to labor, and the initial entry of young adults into the workforce. Each source of change in workforce skill-composition corresponds to an actual flow of workers in the economy. Some of these flows are intra-workforce in character—that is, the workforce composition changes in response to movements of labor workers to professional and pro-

workers experience increasing difficulty in finding employment, more are willing to forego professional status and enter the labor job market. On the other hand, countervailing pressures (such as the availability of transfer payments) may keep the unemployed professional from entering the labor job market as rapidly as supply and demand conditions would alone dictate.

The entry of new workers into the workforce also impacts the mix of laborers and professionals. Figure 1 identifies three distinct flows which alter workforce composition in response to new workers entering the workforce: immigrations, movements of farm workers to labor, and young adults joining the paid workforce for the first time. New immigrants almost always enter the labor category. Even when immigrants possess professional-level skills, they are often prohibited from applying these skills due to language or ethnic barriers.

To grasp the impact of immigration on workforce composition, consider the following simple example.

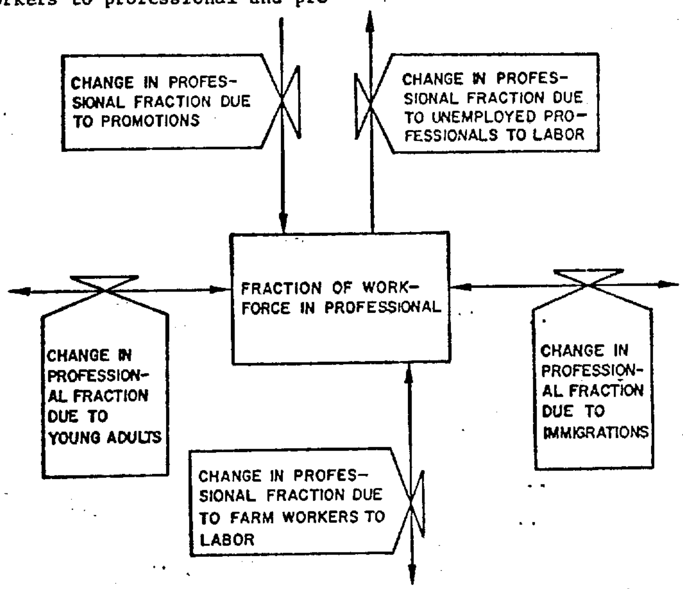


Figure 1. Flows Altering Workforce Composition.

fessional workers to labor. Intra-workforce shifts in workforce composition do not alter the overall size of the workforce. A second subset of worker flows shifts the laborer-professional mix as a consequence of new workers entering the workforce. For example, the arrival of immigrants with labor skills both increases the overall size of the workforce and decreases the fraction of the workforce with professional skills.

The two intra-workforce changes in workforce composition--promotions and movements of unemployed professionals to labor—represent internal shifts in workforce composition in response to relative demand for and supply of laborers and professionals. Experienced laborers move up to fill managerial positions* more rapidly when employers have difficulty filling those positions from the available pool of professionals. Conversely, when an abundance of professionals are available to fill existing professional vacancies, relatively fewer laborers are promoted. The reverse flow of unemployed professionals to seek labor jobs likewise responds primarily to demand and supply conditions. As unemployed professional

If 10,000 new laborers are added to a workforce of 100,000, including 30,000 professionals, the new input of laborers will alter workforce size and composition. Workforce size will obviously increase to 110,000. Moreover, the fraction of the workforce with professional skills will fall to 30,000/110,000, or 27%, as a consequence of adding new laborers but no new professionals to the workforce.

The flow of unpaid farm workers to labor also increases the workforce size and alters the professional-labor mix in the direction of relatively more laborers. Unpaid farm workers include farmers who were previously self-employed or in some other way received goods rather than a wage in payment for their services. The flow of previously unpaid farm workers to labor includes both the physical migration of farm workers to urban areas and the shift of farm workers from unpaid to paid employment. All farm workers are assumed to possess labor-level skills and therefore enter the labor workforce. As farm workers enter the labor workforce, they alter workforce size and composition in a fashion analogous

*Note that the term "professional" subsumes managers as well as specially trained professionals in the present model.

to the impact of immigrants on workforce size and composition.

The impact of young adults on the labor-professional mix is slightly more complicated than the impacts of immigrants and previously unpaid farm workers. Compared to immigrants and farm workers, a relatively large fraction of adults enter the workforce with professional skills acquired through higher education. Moreover, the impact of young adults entering the workforce on workforce skill-composition varies as the fraction of young adults receiving higher education varies.

A simplified example illustrates the impact of a varying student fraction on the professional fraction. For convenience, let us call the fraction of young adults in higher education the "student fraction" and the fraction of the current workforce with professional skills the "professional fraction." Assume that the fraction of young adults with professional skills (that is, the student fraction) is less than the current fraction of the workforce with professional skills (the professional fraction).* In such a case, the impact of the flow of young adults entering the workforce on workforce composition is to reduce the professional fraction in a manner analogous to the impact of immigrations and the farm-worker flow. However, the impact of the student fraction changes as the student fraction changes. Consider the case of 10,000 young adults entering a workforce of 100,000 workers. If there are initially 30,000 professionals (that is, a professional fraction of .30) and 1,000 of the young adults come from higher education (a student fraction of .10), the professional fraction falls from .30 to .282 (that is, 31,000/110,000). However, if the student fraction rises to .20, the flow of 10,000 young adults into the workforce results in a smaller reduction of the professional fraction: from .30 to .291 (that is, 32,000/110,000). If the student fraction continues to rise until the student and professional fractions are equal, the entering flow of young adults does not alter workforce composition. If the student fraction exceeds the professional fraction, the influx of young adults increases the professional fraction.

The student fraction varies substantially over the national life cycle. Section II.B focuses on the factors assumed to alter the student fraction in the present model.

II.B. Overview of Model Feedback Structure

Figure 2 presents a causal diagram for the workforce composition model. As shown in Figure 2, three principal feedback loops control the mix of laborers and professionals in the national system. The first feedback loop represents the decisions of young adults to seek higher education and the resulting impact of higher education on workforce composition. The second feedback loop shown in Figure 2 describes the impact of promotions on workforce composition and the reverse impact of workforce composition on promotions. The third feedback loop incorporates the impact of movements of unemployed professionals to labor on workforce composition.

Demand for higher education is part of a positive feedback loop affecting workforce composition. As is indicated in Figure 2, an increase in the fraction of the workforce in professional leads to an increase in the demand for higher education, which tends to increase the fraction in higher education and, in turn, leads to further increases in the fraction of the workforce in professional.

Young adults demand higher education on the basis of a variety of factors which reflect their career ambitions and their educational plans. The dependency of demand for higher education on the traditional fraction of the workforce in professional represents the role of family background in determining career ambitions. As it is used in determining demand for higher education, the traditional fraction of the workforce in professional approximates the fraction of professional families in the nation. For example, if the traditional professional fraction equals .20, approximately .20 of all young adults are the children of professionals. Therefore, inclusion of the traditional professional fraction as a determinant of demand for higher education represents the number of young adults drawn to professional employment by virtue of family background.

The fraction of young adults aspiring to professional employment also responds to perceived opportunities for professional employment. The present model assumes that young adults perceive the relative availability of professional openings by observing the ratio of orders outstanding for professionals to total orders outstanding for all workers. The so-called "backlog ratio" can be thought of as the fraction of all job vacancies which are for professional openings. If the vacancies for professional jobs comprise a larger share of all vacancies than professionals comprise of the total workforce, the number of young adults who aspire to professional employment will increase.

Demand for higher education also depends upon the perceived appropriateness of higher education as a means to attain professional employment. One factor which increases the perceived appropriateness of higher education is technological complexity. As technological complexity increases, a larger portion of those apiring to professional employment see higher education as a desirable or necessary path to achieve professional employment.

Lastly, Figure 2 shows that the fraction of young adults who receive higher education depends both on the demand for higher education and on the capacity of the knowledge sector to provide higher education. Undercapacity in the knowledge sector (too few classrooms, too few teachers, excessively high tuition fees) restricts enrollments. Excess capacity can expand enrollments to a certain extent. The relative capacity of higher education, shown in Figure 2, measures under- or overcapacity as a ratio of capacity in the knowledge sector to demand for higher education. When relative capacity is less than unity, the demand for higher education exceeds the capacity of the knowledge sector. When relative capacity exceeds unity, capacity of the knowledge sector exceeds the demand for higher education.**

---

*A fraction in higher education which is less than the prevailing fraction of the workforce with professional skills has characterized US history. In 1970, for example, approximately 38% of the paid workforce were professionals (see O. D. Blau and P. M. Duncan, op. cit.) and approximately 30% of young adults were enrolled in institutions of higher education--four-year colleges and universities, community colleges, and trade colleges. (Source: US Department of Health, Education, and Welfare, Social Indicators, 1973 ed.)

**In the workforce composition model, relative capacity of the knowledge sector impacts the fraction in higher education through the multiplier from capacity of knowledge sector for higher education. The multiplier depends only on relative capacity of the knowledge sector. The figures showing model behavior in Section III below plot the multiplier rather than the capacity of the knowledge sector.

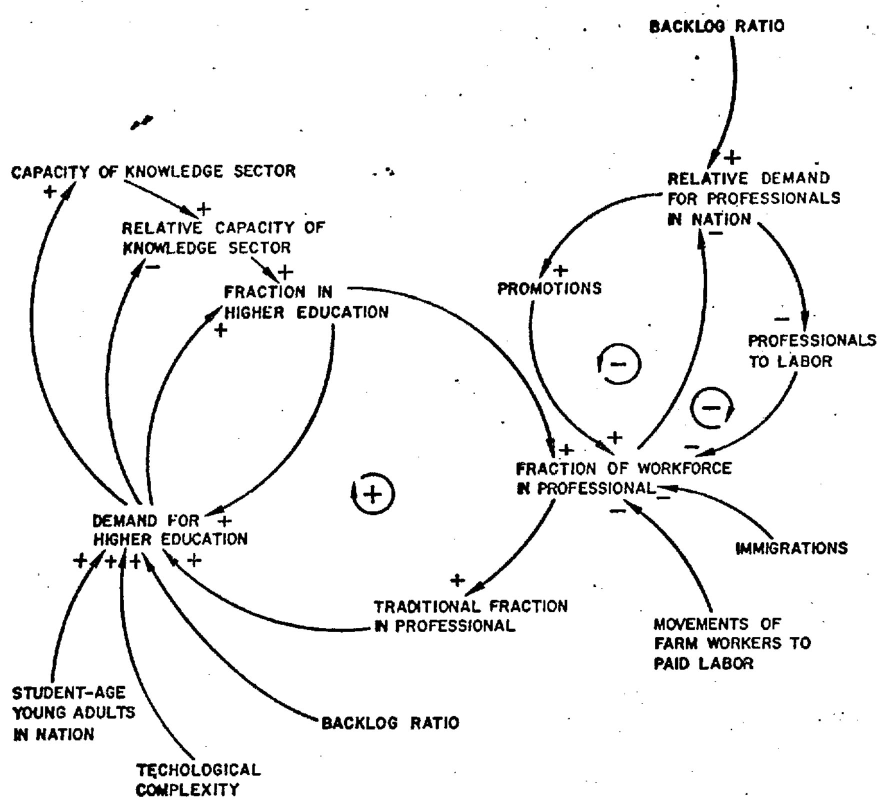


Figure 2. Model Feedback Structure.

In the present model, capacity of the knowledge sector is determined as a twenty-year delay of demand for higher education. The twenty-year delay time in adjusting capacity to demand for higher education represents the lead time necessary to plan for, finance, and achieve additional capacity in the knowledge sector. Representation of capacity as a simple passive adjustment to demand for higher education should be adequate to study interactions of demand and capacity during the period of growth in demand for higher education.*

The higher education loop incorporates three exogenous variables: backlog ratio, technological complexity, and student-age young adults in the nation. The backlog ratio, which measures the relative job vacancies for professionals as a fraction of all vacancies, will be used as a test input in Section III. Technological complexity is input in the present model as an exponentially increasing function of time. The level of technological complexity doubles every 30 years. However, the impact of technological complexity on demand for higher education eventually saturates. As a consequence of technological complexity, the percentage of those aspiring to be professionals who seek higher education as the means to professional employment rises from 58% in 1850 to 98% in 1970. After 1970, further increases in technological complexity only cause the fraction of young adults demanding higher education to asymptotically approach the fraction of young adults aspiring to professional employment.

The last exogenous variable, the student-age young adults in the nation, is generated by an aging chain in the workforce composition model. The aging chain contains five age categories: children (ages 0-14), young adults (ages 15-24), mature adults (ages 25-44), adults beyond childbearing (ages 45-64), and senior adults (age 65 and above). The aging chain can be used to generate

*In the complete national model, capacity of the knowledge sector will be determined in a separate production sector of the model: the knowledge sector will expand in response to orders for higher education as well as to orders for knowledge (technology) from other production sectors in the model. (For an overview of the structure of the generic production sector, see N. J. Mass, "The Production Sector of the National Socio-Economic Model--An Overview," System Dynamics Group memorandum D-2143, MIT) Representation of capacity by an explicit production sector will permit a much broader range of responses in capacity of higher education to changing demands for higher education. Although the present representation of capacity should be adequate for the period of growing demand, representation of capacity as a simple delayed adjustment to demand offers an extremely limited range of dynamic responses when demand levels off.

exponentially increasing population which matches US population from 1850 to approximately 1915 or an S-shaped population pattern which rises to 2.06 million by 1970 and equilibrates at 2.60 million in 2080. The number of student-age young adults is determined as a constant fraction of the young adult population.

The feedback loops involving promotions and movements of unemployed professionals to labor in Figure 2 are both negative loops. Each loop incorporates pressures from the overall economy to adjust workforce composition. These pressures are input through the backlog ratio in the present model. For example, an increase in the promotion rate leads to increases in the professional fraction of the workforce. The increase in the professional fraction alters the relative availability of laborers and professionals and, unless a further increase in the backlog ratio occurs, leads to a future reduction in the promotion rate. The feedback loop controlling the flow of unemployed professionals back to labor responds in an analogous fashion to changing availability of laborers and professionals in the nation.

Finally, Figure 2 shows immigrations and movements of farm workers to labor as exogenous inputs to workforce composition. Immigrations in the US rose steadily from 1850 to 1900 and achieved particularly large values (approximately one million per year) in the years 1900 to 1914.* From that time, immigrations in the US have fallen to a present value of approximately 200,000 per year. The exogenous immigration input in the present model approximates the historical US data from 1850 to 1970. (Source: US Bureau of the Census, Historical Statistics for the US, 1960 ed.) Immigrations are assumed to remain constant at 200,000 from 1970 to 2100.

Movements of previously unpaid (that is, self-supporting) farm workers to labor are likewise input on the basis of historical data. In the case of movements of previously unpaid farm workers, the available data are more imprecise than the immigration data. Data for rural-to-urban migrations from 1850 to the present is used as a surrogate for the movements of farm workers from unpaid to paid labor. Use of such data is inaccurate because some of those who move from farms to the cities are already in the paid workforce and because persons who shift from unpaid to paid agricultural labor are excluded from the rural-to-urban statistic. Data for rural-to-urban migration show a steady increase from 33,000 migrations in 1850 to 270,000 migrations in 1940; since 1940 rural-to-urban migrations fell to a value of 200,000 by 1970. (Source: US Bureau of the Census, Historical Statistics for the US, 1960 ed.) Movements of farm workers to labor is assumed to continue falling to a value of 140,000 in 1990 and remain constant from 1990 to 2100.

III. MODEL BEHAVIOR

Simulation of the workforce composition model presented in Section II shows that the model generates a pattern of growth in the professional fraction of the workforce and the fraction of young adults in higher education similar to that observed historically in the United States. Section III.A discusses behavior of workforce composition in the presence of unlimited exponential population growth. Section III.B investigates the ramifications of equilibration of population for workforce composition and the fraction of young adults in higher education. Section III.C shows how, if given a plausible pattern of behavior, an exogenously determined relative demand for professional and laborers (backlog ratio) can accentuate the basic character of model behavior and bring out an important tradeoff between maintaining a high internal promotion rate and providing higher education opportunities for those who demand them.

*Immigrations to the US fell precipitously during the years of World War I. After the war, immigrations began to rise toward their prewar level until the Immigration Act of 1921 established quotas on immigrations.

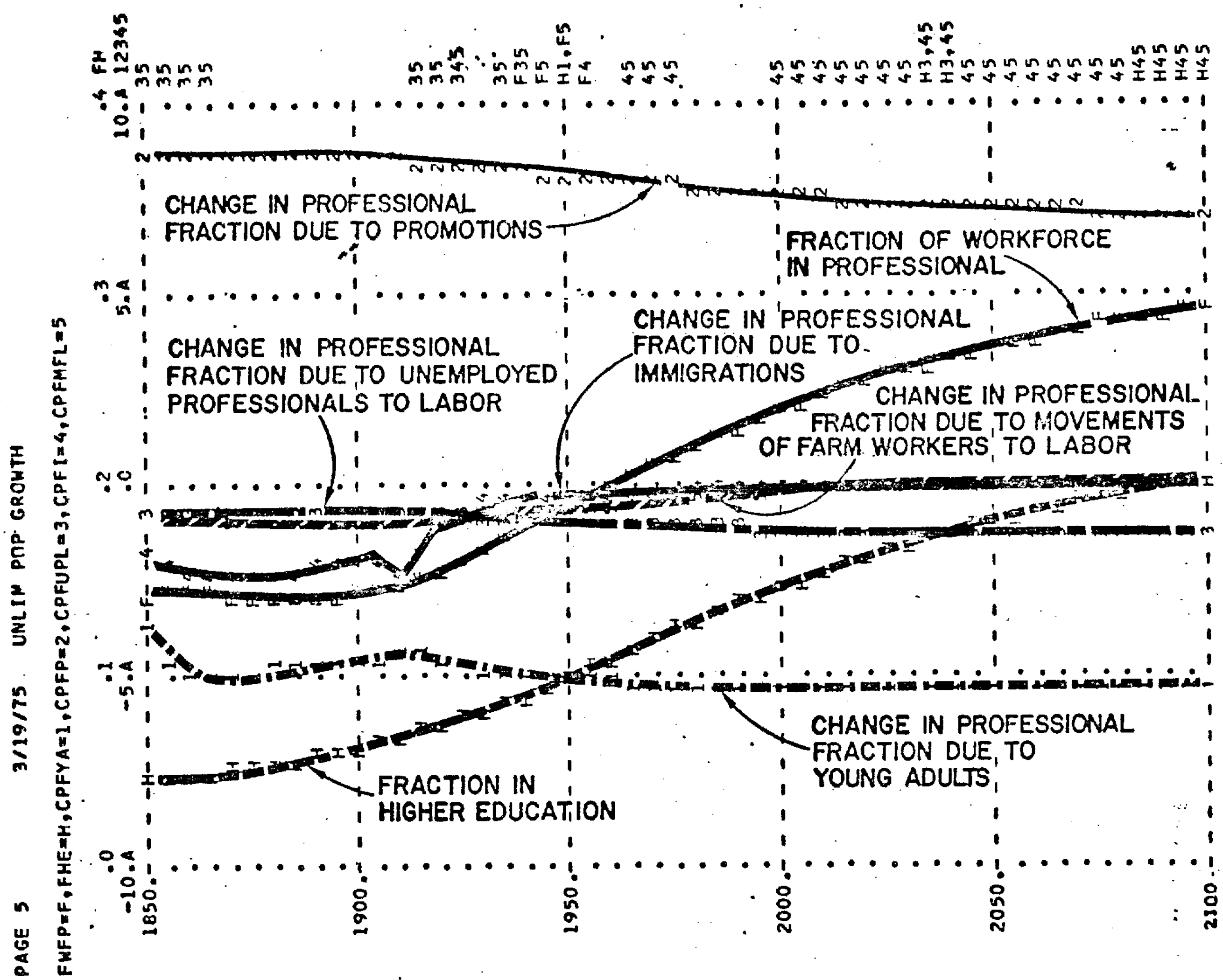


Figure 3(a)

Figure 3. Behavior of Workforce Composition Model Given Unrestricted Population Growth.

III.A. Unrestricted Population Growth

Figure 3 shows the behavior of the workforce composition model when driven by an exponentially increasing population. Figure 3(a) shows the fraction of workforce in professional, the fraction (of student-age young adults) in higher education, and the five rates which alter the professional fraction. Figure 3(b) shows the behavior of the exogenously-determined population (the sum of the five population age categories) and backlog ratio which drive the workforce model. Figure 3(b) also presents a variety of variables involved in the feedback structure which determines the fraction in higher education.

The simulation in Figure 3 shows two major checks on growth in the professional fraction--a large inflow of immigrants with labor skills and a low fraction in higher education. Little growth occurs in the simulated professional fraction between 1850 and 1900 as both these checks combine to offset a large promotion rate. Promotions contribute the only positive rate impacting the professional fraction. In addition to the large negative rates influencing the professional fraction due to the influx of young adults and immigrations, smaller negative rates due to the flow of unemployed professionals back to labor and the movements of farm workers to labor offset promotions.

The balance between promotions, on the one hand, and immigrations and entering young adults, on the other hand, lasts until about 1910. The rapid reduction in the change in professional fraction due to immigrations which begins in 1915 removes a major check on growth in the professional fraction and initiates a prolonged period of changing workforce composition. As immigrations

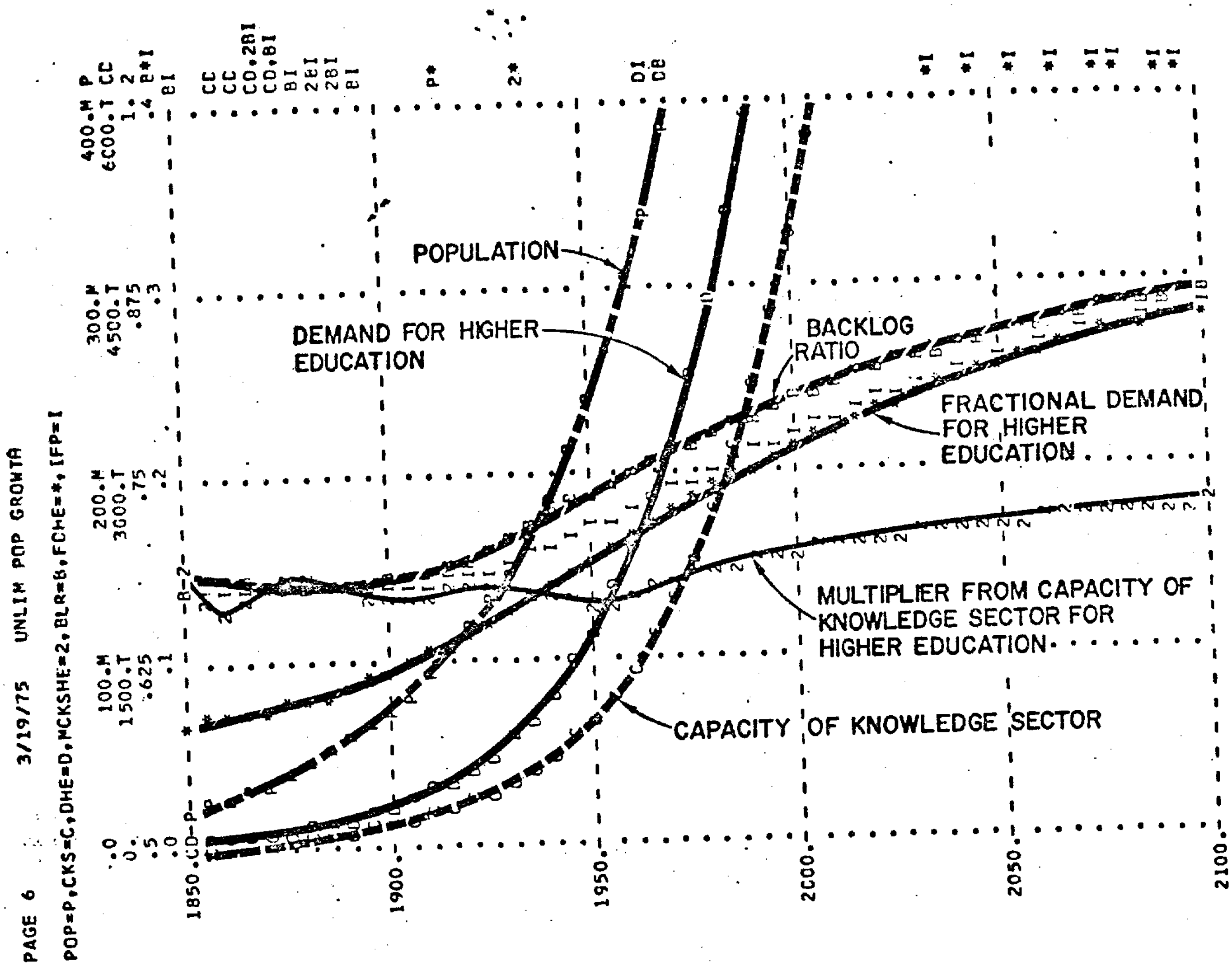


Figure 3(b)

fall, the internal promotion rate pushes the professional fraction upward. Prior to 1910, the large influx of immigrant laborers offsets a high promotion rate; after 1910 the reduced flow of immigrants permits the professional fraction to rise steadily.

The elimination of immigrations as an effective check on growth in the simulated professional fraction occurs as a result of a reduction in the immigration rate which drives the present model. As was noted in Section II.B, an historical reduction in the flow of immigrants into the US occurred in 1915 (see preceding footnote).*

Although immigrating laborers cease to be a major check on growth in the professional fraction in Figure 3, the influx of young adults continues to be a strong force preventing workforce composition from changing too rapidly. As can be seen in Figure 3(a), the change in professional fraction due to young adults remains a large negative rate throughout the simulated period of study. The influx of young adults continues to restrain growth in the professional fraction because there are still relatively few young adults who enter the workforce with professional skills. Even though the fraction in higher education rises steadily over the period 1850 to 2100 in Figure 3(a), the fraction in higher education remains considerably below the fraction of the workforce in professional. Because the rate of change in the professional fraction due to young adults depends on the discrepancy between these two fractions (the fraction of the workforce in professional and the fraction in higher education), the rate remains fairly steady over the entire simulated period.

*When immigration rate is determined endogenously in the complete national model, the impact of falling immigration rate will undoubtedly be more gradual. Nevertheless, the impact of changing immigration rate on workforce composition in the complete model should be similar to that discussed in this section.

The simulation shown in Figure 3 assumes an exponentially increasing population and a ratio of orders for professionals to orders for all workers (backlog ratio) which equals the current fraction of professionals in the workforce. Sections III.B and III.C investigate the ramifications of more realistic patterns of population growth and demand for professionals and labor.

III.B. Population Equilibrium

Figure 4 shows that population equilibrium accelerates and prolongs the growth in the fraction of the workforce in professional. The population input shown in Figure 4(b) closely matches historical population in the US to the present, and equilibrates at 260 million in 2050. As was the case in Figure 3, the fraction of the workforce in professional (Figure 4(a)) remains approximately constant until the flow of immigrants diminishes in 1915. However, whereas the professional fraction rises to .29 at year 2100 in Figure 3, the professional fraction rises to .39 at year 2100 in Figure 4.

Comparison to Figure 3 reveals that the professional fraction rises more rapidly in Figure 4 due to increases in the fraction in higher education brought about by population equilibrium. In effect, population equilibrium removes the second major check on growth in the professional fraction. Unlike the unrestricted growth case, the discrepancy between the fraction of the workforce in professional and the fraction in higher education continuously diminishes in Figure 4(a), leading to a continuous reduction of the change in professional fraction due to young adults. Consequently, the change in professional fraction due to promotions exceeds the sum of the four negative rates impacting the professional fraction throughout the period 1850 to 2100.

The fraction in higher education rises more rapidly in Figure 4 than in Figure 3 because the slowdown in population growth allows growth in capacity of the knowledge sector to "catch up" to growth in the demand for higher education (Figure 4(b)). Beginning around 1915, population (Figure 4(b)) begins its gradual approach toward equilibrium. At the same time, the rate of growth of demand for higher education likewise begins to slow. (Demand for higher education is measured in people and is formed by combining the fractional demand for higher

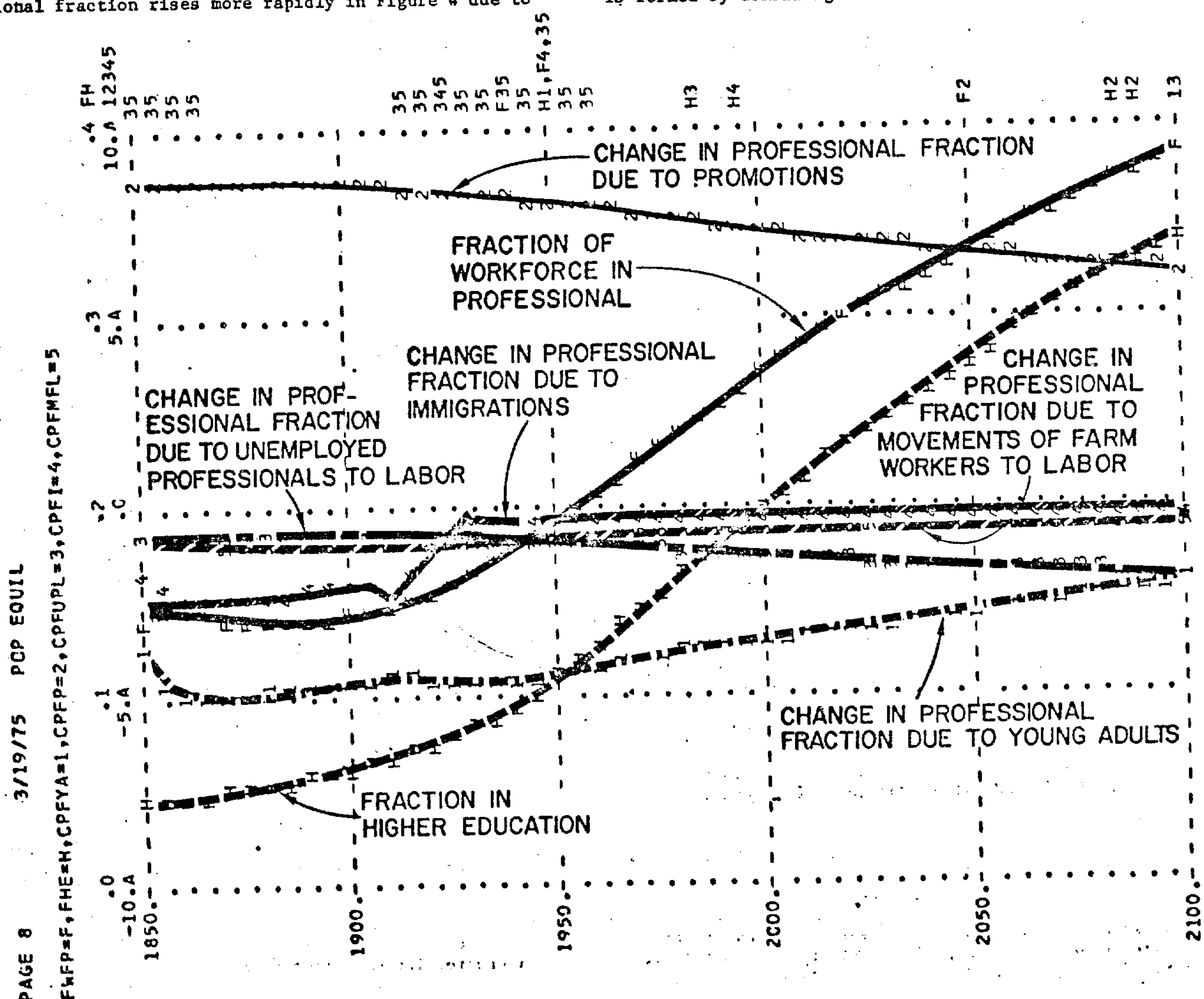


Figure 4(a)

Figure 4. Behavior of Workforce Composition Model Given Population Growth and Equilibration.

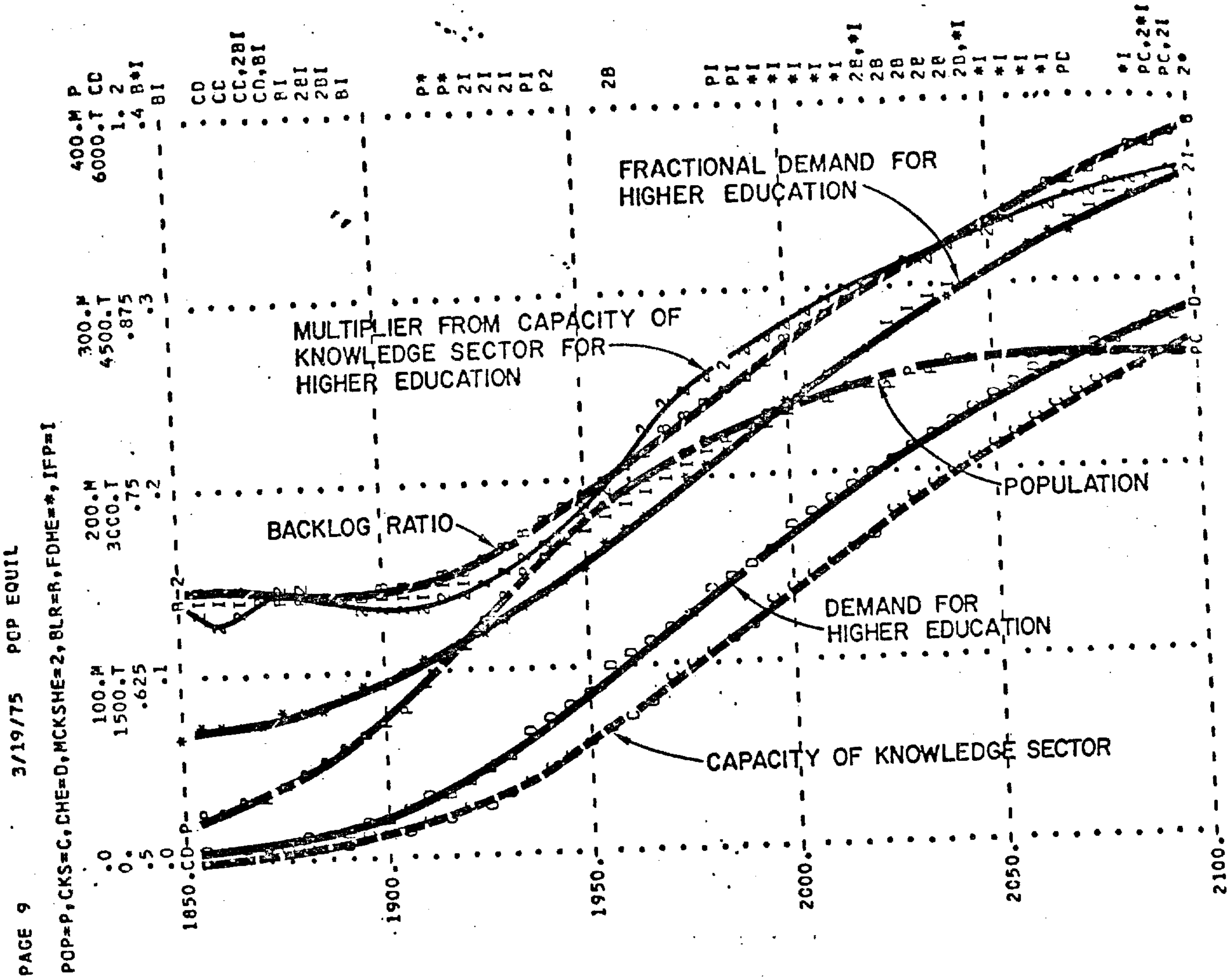


Figure 4(b)

education and the number of student-age young adults.) Capacity of the knowledge sector (likewise measured in people) begins to "catch up" to demand for higher education once demand stops rising exponentially. This occurs in the model because capacity is formulated as a delay of demand for higher education. In real life, such a response would occur because the current rate of growth of capacity of higher education (buildings, teachers, educational materials) depends upon investment and planning decisions made on the basis of past rates of growth in demand. Thus, once the current rate of growth in demand slows, the rate of growth in capacity continues to increase for some time.

Because both demand for higher education and capacity of the knowledge sector are increasing steadily, the "catching up" of capacity to demand is difficult to perceive visually in Figure 4(b). However, the shift is readily apparent in the multiplier from capacity of knowledge sector for higher education. The multiplier, which depends on the ratio of capacity of knowledge sector to demand for higher education (called the relative capacity of knowledge sector in Section II.B), increases from about .67 in 1915 to about .95 by year 2100 in Figure 4. By contrast, the multiplier from capacity of the knowledge sector for higher education rises only to .72 when population growth is unlimited (Figure 3(b)).

The sharp increase in the availability of higher education (signified by the rise in the multiplier from capacity of knowledge sector for higher education) means that more of the young adults who seek higher education can receive higher education. More of those who seek higher education can recieve it because there are relatively fewer young adults competing for available capacity of the knowledge sector once demand stops rising exponentially. The effect of the increased availability of higher education is a much more rapid rise in the fraction of young adults in higher education (the fraction in higher education rises to about .35 by 2100 in Figure 4(a), as compared to .22 in Figure 3(a)) and, consequently, a greater increase in the professional fraction of the workforce over the simulated period from 1850 to 2100.

III.C. Response of Workforce Composition to an Exogenous Demand for Laborers and Professionals in the Nation

In Sections III.A and III.B the relative demand for professionals and labor, represented by the backlog ratio, equals the current fraction of the workforce in professional throughout the simulations. Such an assumption implies that the productive sectors of the economy can absorb any ratio of professionals to labor with no impact on promotions, unemployed professionals to labor, or the other rates which affect workforce composition. In order to increase the realism of the analysis, Section III.C tests the response of the workforce composition model to an external demand for a larger fraction of professionals from 1850 to 1950 than are available and for a constant fraction of professionals after 1950.

Figure 5 shows the response of the workforce composition model to a ramp input in the relative demand for professionals and labor. The exogenous backlog ratio begins at a value of .20 in 1850 (compared to a professional fraction of .14) and rises steadily until a value of .35 is reached in 1950. Backlog ratio remains at .35 for the period 1950-2100

As can be seen in Figure 5, the ramp input in backlog ratio means that, for the first one hundred fifty years, a smaller fraction in professional exists than is being ordered. During this period, production sectors are placing a larger fraction of orders for professionals than the economy can provide. Consequently, the professional fraction and the fraction in higher education grow more rapidly in Figure 5 than in Figure 4. After the year 2000, the professional fraction exceeds the backlog ratio and a larger fraction of professionals exists than is being ordered. Once the professional fraction exceeds the fraction of professionals being ordered, the professional fraction adjusts to an equilibrium value more quickly than occurs in the preceding simulations. A comparison of the professional fraction in Figures 4(a) and 5(a) shows that equilibrium has not yet been reached by the year 2100 in Figure 4(a), whereas the professional fraction stabilizes by about 2050 in Figure 5(a).

Figure 5 shows that the balance of forces that eventually stabilize the simulated professional fraction

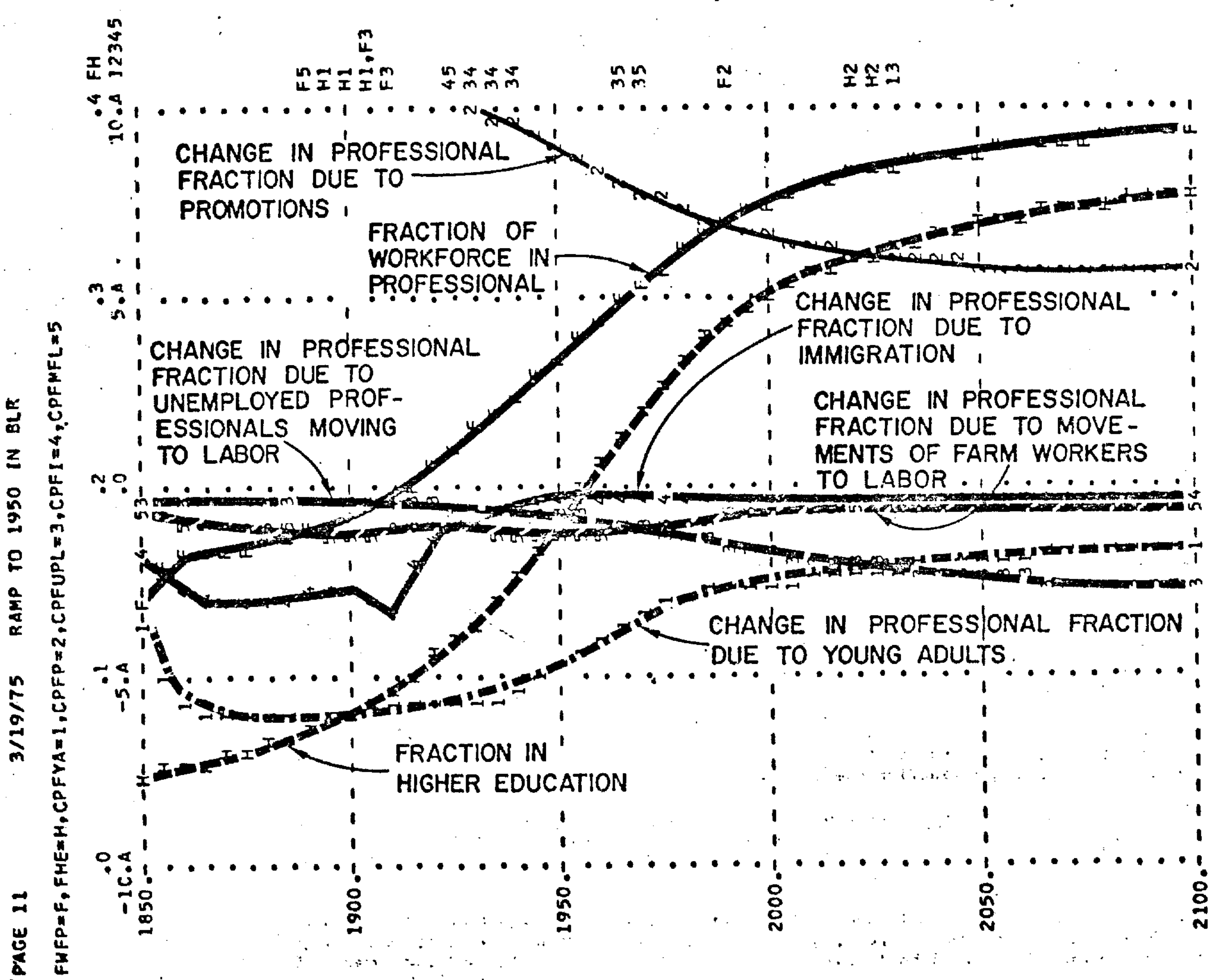


Figure 5(a)

Figure 5. Behavior of Workforce Composition Model Given an Exogenous Input for Demand for Laborers and Professionals in the Nation.

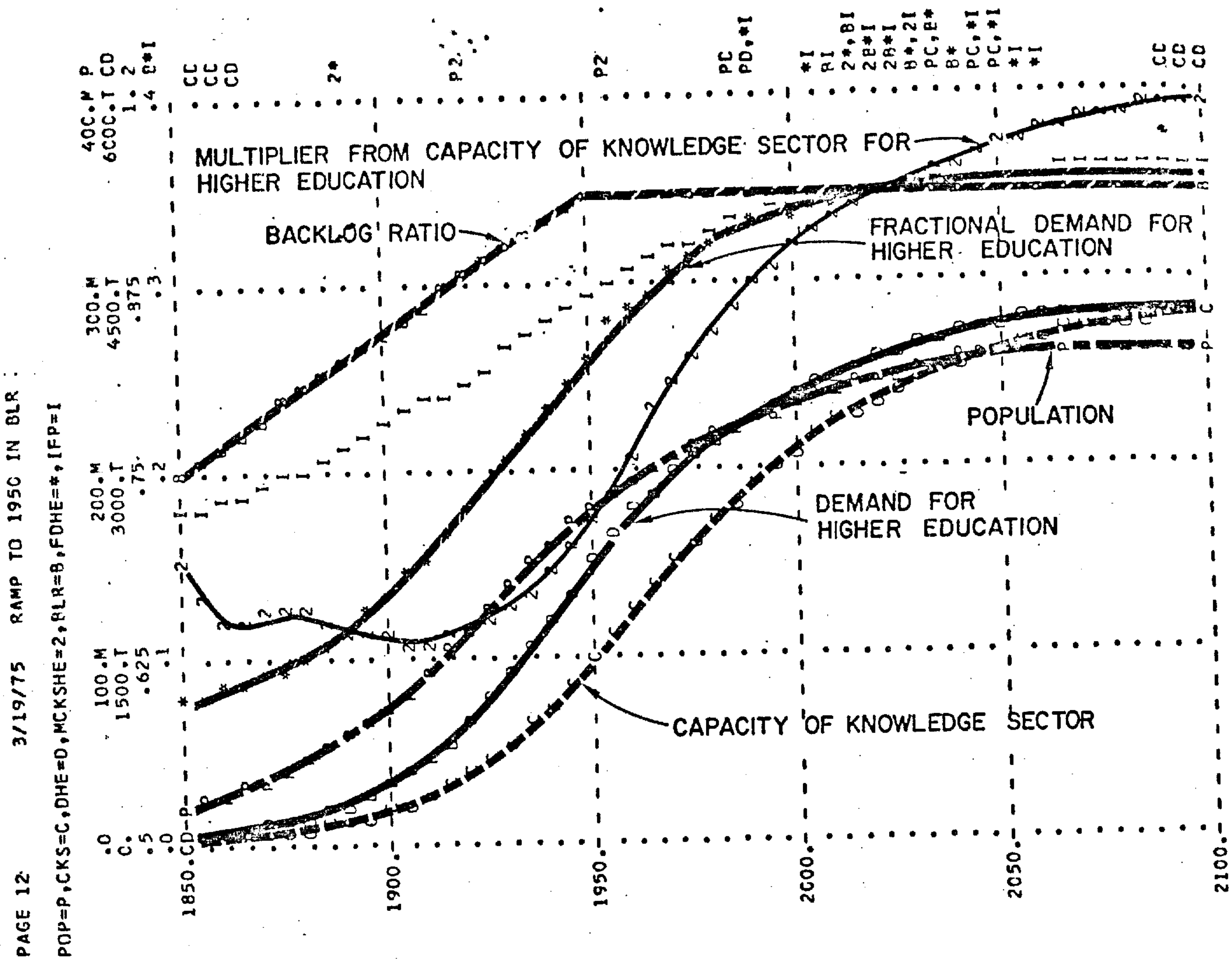


Figure 5(b)

at the end of the growth cycle is fundamentally different from the balance of forces which characterize the stable workforce composition during the early stages of growth. During the period 1850 to 1900 workforce composition in the model system remains fairly constant as a large upward flow of laborers to professional (promotions) is balanced by large influxes of immigrants and young adult laborers. When a stable workforce composition is achieved once again around year 2050, a much lower upward economic flow due to promotions exists. The lower promotion rate is counterbalanced by (1) a change in professional fraction due to young adults which is much smaller than existed during the early stages of growth and (2) by a reverse flow of unemployed professionals to labor that is much larger than existed during the growth phase. If the initial period of stable workforce composition can be called a period of high upward mobility and large influxes of new laborers, the final steady state condition is one of low upward mobility, small influxes of new laborers, and high downward (that is, professional-to-labor) mobility.

The simulation in Figure 5 demonstrates that, in the absence of continuous influxes of new laborers, a stable workforce composition can only be achieved by creating a relatively low promotion rate and a relatively high flow of professionals back to labor. As was discussed in Sections III.A and III.B, a high promotion rate can coexist with a stable workforce composition provided large numbers of new laborers are continuously injected into the economy. In the absence of large influxes of immigrant laborers or a small fraction of young adults who enter the professional workforce, no effective counterpressures to a high promotion rate exist. Therefore, stabilization of workforce composition requires that promotions fall to a rate that can be counterbalanced by the reverse flow of professionals back to labor.

The workforce composition model therefore illustrates a basic tradeoff between maintaining a high rate of upward mobility for the individual laborer and providing professional training through higher education to large fractions of young adults. The simulation in Figure 5 leads to a condition of high fraction in higher education and low upward mobility. Alternatively, if capacity of the knowledge sector were constrained, workforce composition could stabilize at a higher rate of upward mobility and a lower student fraction. Still another alternative mix of laborer mobility and enrollment in higher education could be achieved by constraining many graduates of higher education to assume labor positions upon entry into the paid workforce.

IV. SUMMARY

The model of workforce skill-composition presented in Sections II and III suggests that there exist two major checks on long-term shifts in the mix of professionals and laborers in a national system. Both checks take the form of large influxes of laborers who enter the workforce during growth. To understand how these inflows establish checks on the workforce composition, the model first describes an initial period during which little change in workforce composition occurs.

Throughout the early stages of national growth, the size of the paid workforce continuously expands due to immigrations and due to the influx of ever-increasing numbers of young adults seeking paid employment for the first time. Both flows of new workers affect the laborer-professional mix. The inflow of immigrants is comprised almost entirely of laborers and therefore tends to depress the fraction of the workforce possessing professional skills. During the early stages of growth, the influx of young adults is likewise comprised mostly of individuals possessing labor-class skills. In the face of these large inflows of new laborers, a high rate of promotions of experienced laborers to professionals is required to maintain a more or less constant mix of professionals and laborers in the economy. When applied to the United States, the workforce composition model presented in Sections II and III shows a fairly constant professional fraction (equal to approximately .15) for the period 1850 to 1910.

Once the flow of labor immigrants is reduced in the model system, the professional fraction of the workforce begins to expand. The reduction in immigrations, which corresponds to the drop-off in immigrations in the United States around 1915, leads the simulated promotion rate to dominate the impact of immigrations and the influx of young adults on workforce composition. The discrepancy between promotions, on the one hand, and the factors tending to reduce the professional fraction, on the other, results in prolonged expansion of the professional fraction of the workforce.

Population equilibration removes the second check on growth in the professional fraction of the workforce --the low fraction of young adults who enter the workforce with professional skills. As the rate of population growth slows in the model, the capacity of the knowledge sector draws closer to the demand for higher education (measured by the number of young adults seeking higher education). This allows a larger fraction of young adults to receive higher education and, in turn, stimulates growth in the professional fraction of the workforce. As the fraction of young adults in higher education rises toward the professional fraction, the labor-professional mix shifts due to the promotions of laborers to professionals.

Once the checks on growth in the professional fraction are removed, simulated workforce composition changes until a new balance of factors trying to increase and decrease the professional fraction of the workforce is attained. The conditions for the stable workforce composition which finally emerges are sharply distinguished from the conditions prevailing during the early stages of national growth. In the simulations discussed in Section III.C, the fraction of the workforce in professional stabilizes at about .38 in the year 2050. Workforce composition stabilizes primarily as a result of a reduction in promotions of laborers to professionals and an increased flow of unemployed professionals back to labor. The conditions for stabilization of workforce composition differ markedly from the high upward economic mobility characteristic of the early stages of national growth.

ISSUES UNDERLYING THE REPRESENTATION OF SOCIAL VARIABLES IN SYSTEM DYNAMICS MODELS

Dale Runge
System Dynamics Group
Alfred P. Sloan School of Management
Massachusetts Institute of Technology, Cambridge, Massachusetts

INTRODUCTION AND SUMMARY

Social system simulation models are probably most commonly criticized for "oversimplification." While regarded as useful for analyzing physical systems, mathematical models are often considered woefully inadequate for capturing the complexity and subtlety of human behavior. Such criticisms are based on the existence and importance in social systems of such intangibles as feelings, beliefs, attitudes, and values. This paper aims to explore some basic issues surrounding the representation of intangible social variables in simulation models. Two sources of opposition to the practice of including social variables in simulation models will first be examined and evaluated. The reasons why in principle social variables can be treated the same as physical variables in system dynamics models are then discussed. Finally, the influence of model time horizon on whether a variable is treated explicitly or implicitly in modeling practice is examined.

To facilitate discussion, a convenient way should be available for referring to concepts that most people agree are amenable to mathematization as opposed to concepts whose mathematization is widely considered impossible or inappropriate. The dichotomy "tangible" vs. "intangible" is not adequate for this purpose, since, although all feelings and beliefs are intangible, so also are the well-mathematized concepts of force, energy, and pressure. Force and energy are intangible, and yet are recognized to behave in ways that can be explained and predicted through mathematics. The dichotomy "material" vs. "cultural" suffers from a similar weakness. Although the term "cultural" connotes people and their attitudes and actions, the term "material" is synonymous with "tangible," and therefore has limited applicability. In the absence of a more flexible dichotomy, this paper will distinguish "physical" from "social" variables. Although "physical" tends to imply "tangible," physical <u>systems</u> are commonly thought of as a well-defined class of systems. As a result, all the variables in physical systems, tangible or not, will be considered physical variables. Physical variables also include tangible components of social systems, such as capital goods, inventory, and population. For the intangibles of social systems, such as preferences, goals, desires, beliefs, attitudes, and values, the term "social" variables should suffice. Any intangible concepts that are unique to social systems fall into the category of "social" variables.

The criticism referred to above about "oversimplification" in mathematical models has been especially evident in many critiques of the recent system dynamics-based books <u>World Dynamics</u> and <u>The Limits to Growth</u>. The authors of <u>Limits</u>, for example, allegedly "ignore the real world of social, political, and ethical values," and are criticized because "one of the most obvious features of human society is that values are constantly changing and values affect behavior."[1]* According to another observer, the books "hardly concern [themselves] with the rates of change in human habits and institutions, and [they] ignore almost all of politics, economics, and the other social and behavioral sciences."[2]

Some of the issues raised in the above comments naturally relate to how the specific models either included or failed to include value change. More basic issues involved in any such criticism must also be addressed to provide a context for discussing the representation of social variables in specific system dynamics models. Perhaps the fundamental issue is the question whether in principle the modeler can capture subjective social variables and forces causing their change in terms of mathematical models. A related question is whether social variables for which little or no data exist should be included in mathematical models. These two issues will be discussed in turn below.

THE POSSIBILITY OF MATHEMATIZING SOCIAL VARIABLES

Much of the criticism of the representation of social variables in mathematical models seems to spring from disagreement over the applicability of the scientific method to the study of social systems. Whether the generation and testing of theories can lead to explanation and prediction in social systems has been a source of contention for many years. Recently, historians have joined in vigorous dispute over the applicability of the deductive methods of the physical sciences to the study of human society. While much of the dispute revolves around such philosophical issues as the nature of explanation itself,[3] a large part of the concern within history and other social disciplines hinges on the use of mathematics to describe social processes. For example, historian Stuart Bruchey argues: "to force a translation of qualitative factors into the language of numbers is to guide us to reality in the way of parody."[4] But what is it about "the language of numbers," or mathematics, that stimulates such a criticism? Indeed, the precision and consistency of mathematics are largely responsible for the power of explanation scientists have of physical phenomena.

Apparently, the very precision required in formulating a mathematical model is itself a source of resistance against the practice. The ambiguity with which everyday language describes experiences, feeling, and emotions appears related to our own experiences of the subjective concepts being described.[5] Increased precision in describing subjective matters means a greater gap between the description itself and what is being described. For example, the mathematization of attitudes towards welfare payments in no way conveys the experience of having those attitudes. The distinction between a description of an experience and the experience itself appears to be missing in many reactions to the objectification entailed by using mathematics to describe subjective states of being. According to Rudner,

> The alleged failure of social science to "capture" (i.e., to reproduce or to be the psychological equivalent of) the delighted chortle of a baby in social play with its parent, the anguished embarrassment of an adolescent, the nuances of social interaction of a board of directors meeting or of a cocktail party, is too often nothing but the <u>failure to distinguish</u> statements and systemizing uses to which they

*Superscripts refer to notes at end of paper.

may be put, from the social phenomena referred to by those statements."[6] [underline added]

When the "statements and the systemizing uses to which they may be put" are mathematical in nature, the description of the experience is even further removed from the experience itself; hence, a mistaken tendency to reject the validity of the description. But a mathematical model is not supposed to reproduce the described experience; it merely aims to describe or even reproduce the processes by which a particular subjective state arises and the impact of that state on the rest of the model. The main difference between a verbal and a mathematical description is the greater precision of the latter. The increased precision of a mathematical description does not per se mean that it cannot deal with social concepts; there is no barrier in principle to treating social variables the same as physical variables in mathematical models. A later section will explore in some detail the similarity between physical and social variables from the viewpoint of system dynamics methodology.

APPROPRIATE SOURCES OF INFORMATION FOR MODELS

Even those who share the assumption of the applicability of the scientific method to social systems disagree over whether variables which are not quantified, or relations between variables that cannot be tested by statistical means, should be included in mathematical models. One end of the spectrum of opinion is found in econometrics, where the focus of attention is on testing hypothesized relationships between time-series data of measurable economic variables. Moreover, strict disciplinary bounds seem to preclude the use of variables from fields other than economics. Persistent efforts in the field of "behavioral economics" seem to have made little headway in introducing new variables into traditional econometrics. For example, the effect of consumer attitudes on economic behavior are claimed by econometricians to be better explained by economic variables than by attitudinal data.[7] Reports of intended investment by businesses are found to be incorrect after the fact.[8] Such instances seem to validate the economist's concentration on quantified variables in explaining economic behavior. Criticisms of the concentration on quantified variables can be found within the field, however. According to Simon Kuznets, "concentration on quantifiable factors in formulating hypotheses may mean a definite bias in the selection and too high a price for a statistically testable hypothesis."[9] The confinement of traditional economics to quantifiable factors has also been criticized by E. H. Phelps Brown in a presidential address to the Royal Economic Society:

> ...[M]y argument implies the removal of the traditional boundary between the subject-matter of economics and other social sciences....
>
> ...When the actual way in which decisions are reached in the board room or across the bargaining table has been discussed, it has been said that economics as such has nothing to contribute. Down with "economics as such."[10]

The viewpoint within system dynamics lies at the opposite end of the spectrum of attitudes towards appropriate sources of information for mathematical models. System dynamics shares the criticisms of confinement to disciplinary boundary, and of exclusive consideration of variables for which time-series data exist. For a system dynamics model, the variables and their interactions are specified according to their contribution to the dynamic behavior of the system in question, not according to any disciplinary criteria. A system dynamics model is a hypothesized set of interactions that produces the behavior of concern. The modeler is obliged to include all the inputs to decision processes that can be expected to influence model behavior significantly. In modeling economic development, for example, this approach is especially important, since by widespread consensus, cultural, political, and social variables as well as economic variables are influential.

Including the effects of such forces does not necessarily mean that the forces themselves must be made explicit, however. Critics of World Dynamics, for example, often claim that attitudes and values are not accounted for in the model. In general, variables which adjust rapidly relative to the time horizon of a model can be treated implicitly rather than explicitly. Consequently, the price of natural resources--a social variable--is not explicitly represented in the World Dynamics model. As resources become scarce, higher price discourages their use. Price is an intervening variable between availability and usage, and since price is merely a way of linking the two, an explicit treatment of price is not necessary. The same comment applies to the social values that impinge on birth rate. Instead of an explicit treatment of the relevant values and the effects on those values of crowding, material standard of living, food, and pollution, the model links those physical factors directly to the birth rate equation. The values are by no means "ignored"; an explicit treatment of them was simply not necessary to express their influence on the behavior of the model.

SYSTEM DYNAMICS METHODOLOGY AND SOCIAL VARIABLES

System dynamics is a methodology for understanding the dynamic behavior of systems. A system dynamics model contributes to a better understanding of the causes and processes of change, and provides leverage for altering the behavior of the system under study. Besides providing a set of techniques for modeling, system dynamics employs a set of views about the nature of real-world systems. This section of the paper discusses some of those views to establish the suitability of the system dynamics methodology for treating social variables.

Any mathematical model attempts to organize information about real-world relationships and processes. The method of organization varies widely, however, from attempts to discern correlations to attempts to establish causal relations among variables. A system dynamics model is more than a collection of equations that reproduce the time paths of system variables. System dynamics models are meant to replicate system processes, as opposed to the outcomes of those processes. System dynamics models are commonly characterized as causal models, in contrast to other types of models that express correlations between the time behavior of variables. A causal model necessarily deals with the rates of change of system variables over time. Therefore, a causal model can be thought of as a set of differential equations. System dynamics models can be described as sets of coupled non-linear differential equations.

Another characterization of system dynamics models appears more suited to the purposes of this paper, however. Integration is the basic dynamic process which a system dynamics model simulates. The time behavior of a system dynamics model is generated by integrating the differential equations that represent the causal relations in the model. A principal tenet of the methodology is that the process of integration in the model is the same as integration in the real world. In fact, of the two mathematical processes of differentiation and integration, J. W. Forrester asserts that only integration occurs in the real world.[11] The symbols used in diagraming system dynamics models reflect the emphasis on integration. The occurrence of integration is represented by a box symbol ☐ called a <u>level</u>. The symbol and name are intended to draw attention to the process of integration or accumulation associated with, for example, the phrases "level of water in the glass" or "level of population." The level variable can be increased or decreased, as, for example, when the level of population is increased by birth and decreased by deaths. The valve symbol ⊏⋈ called a <u>rate</u> represents actions affecting the level variable. The rate symbol intentionally connotes flow or change, while the phrase "rate of change" readily suggests the actions of production, consumption, or change of attitude. A fundamental principle of system dynamics is that these two symbols--and the processes they represent--are sufficient to represent the dynamics of any system.

System dynamics also stresses the phenomenon of feedback. Feedback includes the effect that the position of a body has on its own subsequent motion, the effect that the size of a population has on birth rate, and the effect that information about the size of a firm's inventory has on its production of goods. Feedback of information about the value of level variables--the state of the system--to the rates affecting the levels gives coherence to a system. Feedback gives meaning to the term "system." Without feedback, the idea of a system would imply no meaningful, persistent relations. Given feedback, a boundary can be drawn around a system and the processes and relationships endogenous to the system (as defined by the boundary) can be analyzed. Within the boundary chosen and over the time horizon of interest, the process of integration and the feedback of information are sufficient to simulate the dynamics of <u>any</u> system. This generalization applies whether the components of the system are physical, economic, demographic, social, or psychological. The emphasis here is on the <u>processes</u> that generate system behavior, rather than the context in which the processes occur. Integration and information feedback characterize all dynamic systems, whether the integrated medium is energy, houses, goods, price, or integrity.

The preceding discussion should make clear that the term "model" in the system dynamics sense takes on a particular meaning. System dynamics models were earlier described as replications of the system processes rather than of the outcomes of those processes. A system dynamics model is actually another version of processes occurring in the real world. The integrations and information links in a system dynamics model simulate the processes occurring in the changing real world. From this point of view, no important dynamics distinguish the system dynamics model from the system which the model simulates.

The consonance between mathematical and real-world systems can be seen in models of social systems. For example, consider the interaction between inventory and workforce in a factory. Production and hiring/layoff policies often lead to fluctuations in the levels of inventory and workforce. The basic interaction between inventory and workforce can be summarized as follows: an increase in factory workforce will increase the production rate, and therefore also increase the inventory; if inventory increases above a desired level, the workforce will be reduced to compensate. A very simplified version of these relationships appears in Figure 1 below.

The basic DYNAMO[12] equations for the system in Figure 1 are:

L	INV.K = INV.J + DT*NPR.JK	Inventory (units)
R	NPR.KL = WF.K*WP - S	Net production rate (units/year)
L	WF.K = WF.J + DT*NHR.JK	Workforce (persons)
R	NHR.KL = (DINV - INV.K)/ (TAI*WP*TAWF)	Net hiring rate (persons/year)
C	S =	Sales (units/year)
C	WP =	Worker productivity (units/person-year)
C	DINV =	Desired inventory (units)
C	TAI =	Time to adjust inventory (years)
C	TAWF =	Time to adjust workforce (years)

The equation for net hiring rate NHR states that, whenever inventory is less than desired inventory, more persons

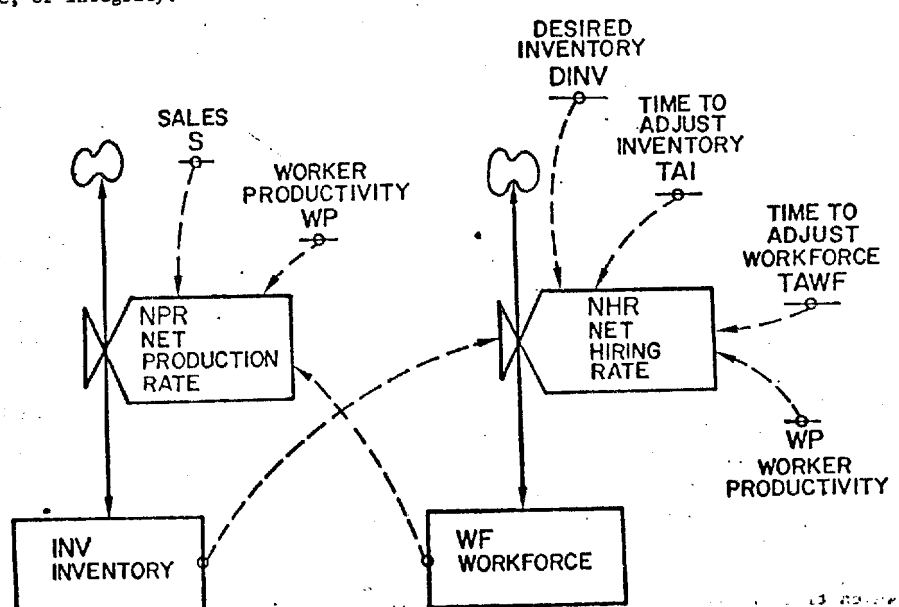


FIG. I

will be hired than laid off, and the size of the workforce will grow. Workforce expands until the discrepancy between desired and actual inventory disappears. The effect of increased workforce on inventory is shown in the equation for net production rate NPR. A larger workforce generates higher net production; as long as workforce grows, so does production. Workforce and production increase until the discrepancy between desired and actual inventory vanishes. At this point, net production will be positive, however, and inventory will continue to grow, exceeding desired inventory. According to the hiring equation, with inventory greater than desired inventory, workers are laid off. The workforce size thereupon begins to contract. As long as inventory exceeds desired inventory, workforce—and consequently net production--declines. The process continues until inventory once again equals desired inventory, at which point the net hiring rate goes to zero. Inventory is falling at this point, however, and once again "overshoots" desired inventory. When inventory falls below desired inventory, hiring must increase, and the entire cycle begins again.[13]

The mathematical system dynamics model of the inventory-workforce system described above produces the same behavior as the real-world system. It does so because the mathematical model contains the same process of integration and feedback, in the same structure relations, as the real-world system. The level variables of inventory and workforce in the example clearly arise from the process of integration. The net accumulation of production less sales creates inventory. Analogously, the workforce is an accumulation of people increased by hiring and decreased by layoffs. The integration that occurs during computation of the level variables of the model reproduces the same process in the real world. The model consists of a mathematical version of structure and behavior in the actual system. Model behavior resembles the real system because the model contains the important real system processes.

In the inventory example, the variables were tangible, and the correspondence between the model process and real-world process was clear. The tangibility of physical variables helps show the importance of integration in processes of change. Inventory is an accumulation of the difference between production and sales; capital stock is an accumulation of the difference between investment and discard; population is the net sum of births and deaths. As the rates which produce accumulations change in magnitude, the values of the levels also change, but the process of integration is unchanging. Beliefs, habits, and feelings can be represented by the same dynamic process, although not so obviously. The intangibility of these concepts prevents any visualization of changes in attitudes and values in the same manner as inventory changes, for example. The notion that intangibles can be modeled in the same way as tangible material variables is therefore commonly rejected. Most observers posit some fundamental distinction between physical and social variables that precludes their respective treatment in parallel ways. The system dynamics methodology, however, recognizes no fundamental distinction between tangible and intangible variables.

The distinction between physical systems, where palpable, countable entities dominate, and social systems, where beliefs, attitudes, and values are prevalent, must be pressed somewhat further. Consider the physical example of a mass suspended from the ceiling by a spring, as shown in Figure 2. The height and speed of the mass and the compression or extension of the spring are all apparently measurable, or at least observable. The equations of motion of the mass deal the the position of the mass and its changes, but the integrations in the system

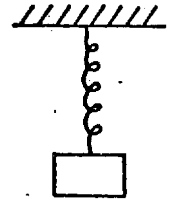


FIG. 2

involve kinetic and potential energies. The energies depend on the position and velocity of the mass; for example, a change in height is synonymous with a change in potential energy. The dynamics of the system, however, are determined not by the height, but by the energy change associated with a change in height. The illusory nature of the "tangibility" of physical systems should now be apparent. "Energy," an intangible physical variable, can only be perceived through its effect on the dynamics of the system.

The important influences on the dynamics of physical systems are less tangible than commonly realized. The regularities of behavior in physical systems allow the concepts of force, mass, and energy to be isolated and interrelated in mathematical models. The regularities of social system behavior, from business cycles to the life cycles of civilizations, suggest similar possibilities for isolation and mathematization of variables that contribute to the dynamics of systems. But a general awareness that social variables involve the same integration process as physical variables is lacking. Consider, for example, the perception of poor quality in a manufactured item, such as razor blades. If customers are used to high-quality blades, but a few poor blades escape detection, customer opinion will start to change. If the frequency of bad blades increases, the customer dissatisfaction will accumulate through a process of integration. If poor quality control persists, dissatisfaction will begin to affect sale of the razor blades. Here customer satisfaction--or its corollary, dissatisfaction--is a dynamic variable, integrating the effects of perceived blade quality to produce a change in magnitude. It takes some time for the appearance of bad blades to be reflected in sales, which is a further sign that integration is occurring in the effect that perception of poor quality has on satisfaction and purchases. All integration processes introduce some delay in a flow channel, whether of information or of matter.

The razor blade example demonstrated the process of integration for a specific social variable. In a more general, long-term way, the integrative nature of social variables is revealed in the following passage:

> Habits and skills accumulate in society as in a reservoir and thus become available to human beings in successive generations.[14]

The cumulative, or integrative, nature of all "habits," including modes of thought, attitudes, and values, should be somewhat clearer from this description. This section has attempted to establish a fundamental similarity between physical and social variables in the context of the system dynamics methodology. As a result of this similarity, there is no barrier in principle to including both kinds of variables in system dynamics models. The contribution to the dynamics of the system that a particular concept or variable--physical or social--is believed to have is the key criterion for selection. To make a selection of variables requires a clear purpose, as well as intimate familiarity with the system being modeled. The next section will illustrate some of the influences that model purpose has on the treatment of social variables.

IMPLICATIONS FOR MODELING PRACTICE

The previous section attempted to establish that, in principle, social variables can be included in a system dynamics model. Physical and social variables can be treated analogously because each involves integration and feedback. However, the selection of variables, physical and social, to include as integrations in a model is determined by the purposes of the model. The behavior modes which the model is intended to reproduce, and the policies which the model will test, govern the composition of a model's variables. The influence of model purpose on model composition is easy to assert, but quite another matter to apply effectively in practice. This section will illustrate the influence of model time horizon (which depends on purpose) on whether a given social variable is treated implicitly as a parameter or explicitly as a dynamic variable--that is, an integration.

Consider the relationship between two variables, A and B, given by the equation

$$A = k \cdot B \qquad (1)$$

Here A and B might represent quantities that change over time, such as inventory, price, razor blade quality, or customer satisfaction. Equation (1) states that when there is a change in variable B, variable A changes immediately to a new value. Let us look closer at this statement. In principle, variable A cannot change immediately in response to a new value of variable B. Some amount of time is required for the information about the new value of B to be transmitted. Some dynamic process must underlie the adjustment of variable A to its new value. The dynamic process underlying Equation (1) might be like that represented below in Figure 3. The rate of change of variable A in Figure 3 is given by:

$$\dot{A} = (k \cdot B - A)/T \qquad (2)$$

If $A = k \cdot B$ in Equation (2), the rate of change of A, or $\dot{A}$, will be zero. Therefore, as long as B remains unchanged, Equation (1) holds true. Given, say, a one-time increase in B, $\dot{A}$ will act to bring A back into equality with $k \cdot B$, at which point $\dot{A}$ will once again be zero. Over some period, however, Equation (1) will not hold true. The length of the period of inequality depends on T, the time constant of adjustment in Equation (2). The time constant in a measure of the speed of adjustment of A to a new value of B. The smaller the time constant, the faster $\dot{A}$ equilibrates A with B, and vice versa.

In other words, an apparently static relationship between variables (provided the relationship is not an identity) actually embodies an assumption about the dynamics of that relationship. Specifically, the time constant of adjustment--relative to other time constants in the system--is so short that the relationship between the two can be <u>approximated</u> as static. In other words, relative to the other dynamic processes in the system, the two variables can be considered to be in equilibrium with each other. The relative lengths of the time constants in a system play a large role in determining the need for an integration to represent a given variable. If the time constant T in Equation (2) is sufficiently short, the value of A at any time can be approximated by knowing the value of variable B and of the parameter k. Even though the value of A can vary, only the values of the parameter k and variable B must be known to determine A. A need not be treated as a separate level, or integration.

The particular nature of the modeled system will greatly influence whether variables--physical or social --must be introduced explicitly in the model as integrations, or treated implicitly by parameters. Consider the razor blade example once more. The time constant of change for the level of customer satisfaction determines whether customer satisfaction must be treated explicitly as an integration, or implicitly by using a parameter in the feedback loop between blade quality and rate of

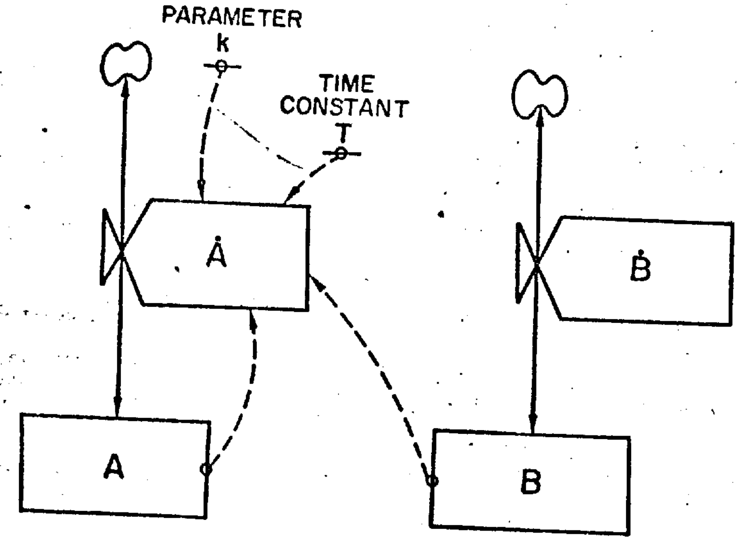


FIG. 3

sales of razor blades.* The relative time constants of the other state variables will determine the choice.

If the time constant of a variable is very short relative to the model time horizon, the variable can be represented implicitly. On the other hand, a variable may change so slowly that it can be approximated as a constant. Holding some variables constant in a model is a useful practice; the dynamic structure needed to generate those variables can be omitted. Therefore, attention can be focused on important elements of the problem that change within the model time frame. To address issues that evolve over longer time horizons, however, more and more constants must be expressed as variables. As the model time horizon is extended, only those relatively basic aspects of the physical or social state of the system can be assumed to be constant. In principle, the modeler should be able to bring parameters into the model as variables as the time horizon extends, until only a fairly small set of basic human values and physical relations need be held constant. At the extreme, human values are likely to resemble basic needs more closely, with very long time constants--that is, which change very little if at all.

The extent to which changes in any variable--physical or social--must be treated in a model depends on the time horizon of the model. For short-term studies of trade cycles, population can be treated as constant. Similarly, resource availability can be assumed constant over short time spans. The same general idea applies to social variables also. Attitudes toward birth control can be assumed constant over a short time span; the public pressure for increasing welfare is constant over short time spans. To omit variables or treat them as constants in a given model structure in no way implies any diminution of the importance of those variables. It merely reflects a pragmatic approximation, based on relative time constants for the changes in variables--always relative to the time horizon of the model. Because the importance of time constants is a relative matter, the example in Figure 3 of the relationship A = k•B can be looked at in another way. If the time horizon of the study is in the range of the time constant T for variable A, attention could focus on the details of the change process of variable A. If variable B changes very slowly over this time horizion, variable B could be treated as a constant in the equation for the rate of change of variable A.

Another illustration should help to clarify this point. We will consider the impact of desired inventory on production rate. In Figure 4, inventory INV should be conceived as a physical integration, such as goods or people. Even though INV is a physical variable, desired inventory DINV is a social variable. The social variable DINV forms part of the policy for the production of the items in INV, as shown in the equation for the production rate:

$$P = (DINV - INV)/TAI \qquad (3)$$

For short time spans, DINV is surely constant, as shown here. Suppose DINV equals 1000 units. This target may be constant over a few days, weeks, or even months, but production policy is unlikely to be independent of sales; if the sales rate varies, as during growth, or over the course of the business cycle, the value of DINV is sure to change. DINV may also change, for example, if the

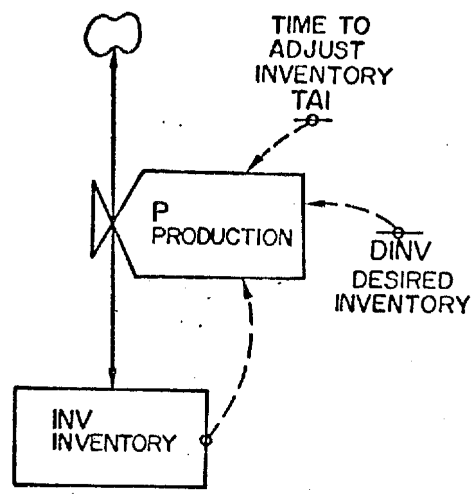


FIG. 4

average sales rate does not change, but the variance in sales increases. If the time horizon for the model is going to include periods of varying sales, the constant DINV must come within the boundary of the model as a variable. The amount of desired inventory usually reflects a desire to meet a constant sales rate for some length of time. In other words, DINV is the number of units in stock required to meet constant sales for a certain period of time if production were zero (See Figure 5).

In Figure 5, average sales AS is an exponential average of the sales rate S. Time is required to average the instantaneous sales rate; consequently, AS is generated by an integration process, and is shown as a level variable in Figure 5. Desired inventory DINV is now a variable, as given by the equation

$$DINV = ICT*AS \qquad (4)$$

Note that Equation (4) has the same form as Equation (1). The form of this equation presumes that the adjustment of DINV to a new value, when AS changes, takes place rapidly relative to the other changes in the system.

In the present formulation, desired inventory has become a variable within the model to account for the effects that variable demand might have on production policies. Therefore, a new parameter, inventory coverage time ICT, was required. The value of DINV is always known if the variable AS and the parameter ICT are known. To take the example a step further, the value of ICT is not a magical number. Why, for example, should a firm hold inventory in the amount of two months of sales rather than three or even ten? The value of inventory coverage itself reflects a balance between the desire to satisfy customers and the costs entailed by carrying high inventory. Company traditions can also be expected to play a significant role. The effect of inventory coverage time ICT on customer satisfaction can be seen in the following way. Equation (4) can be rewritten as ICT = DINV/AS. Here ICT is the ratio of desired inventory to

*Instead of a single parameter, several parameters may be used to express a non-linear relation between blade quality and sales. In either case, customer satisfaction is only implicitly present, since its effect on sales is given by the values of blade quality and of the parameter(s).

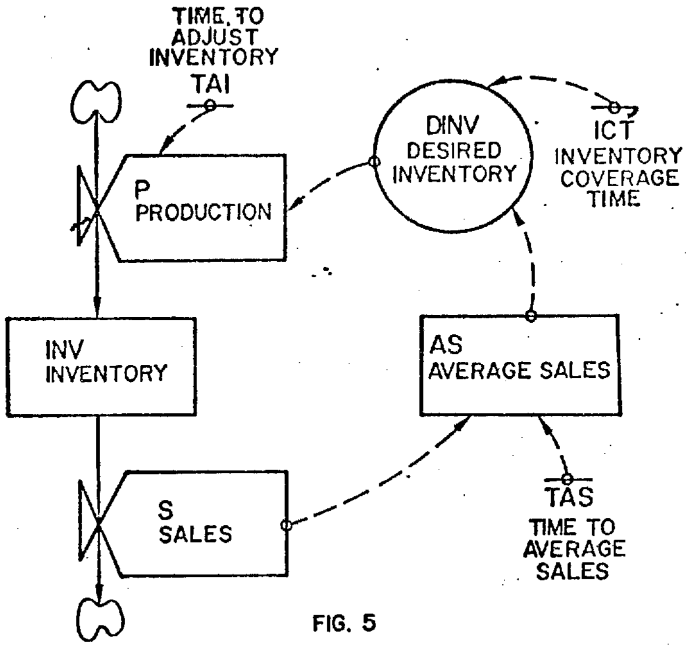


FIG. 5

average sales. The higher the value of ICT, the greater is inventory relative to sales. The ratio of desired inventory to average sales affects the speed with which a given customer order is shipped. Therefore, the value of ICT implicitly measures the fraction of demand that can be satisfied directly from inventory, and thereby affects the average delivery delay to the customer. The effect of ICT on costs is apparent; a higher value of ICT means more inventory on hand at any given time. Carrying inventory costs money; this cost also affects the actual value of ICT. Traditions within the firm influence the extent to which pressure from consumer satisfaction and inventory carrying costs are converted into a change in inventory coverage time. In this manner, inventory coverage time ICT can be brought within the model boundary as a variable. In doing so, other parameters are likely to be necessary; depending as always on model purpose, these parameters can in turn be incorporated into the model as variables. As the time horizon of interest extends, only those variables that are relatively unchanging relevant to the problem can safely be treated as parameters.

CONCLUSIONS

In principle, no barrier constrains the incorporation of social variables in mathematical simulation models. Much of the resistance to doing so arises from the precision with which such models deal with intangible, subjective concepts, and fram a failure to distinguish the description of a concept from the concept itself. Moreover, social variables can and must be included in a model, even in the absence of statistical data for the variables.

Social variables can be treated the same as physical variables in system dynamics models. Both types of variables involve the process of integration and the phenomenon of information feedback. With respect to modeling practice, the purpose of the model governs whether and how both physical and social variables are incorporated. Model time horizon, and the relative time constants for processes within the system, are especially important in determining whether variables are treated explicitly as state variables, implicitly by means of parameters, or are assumed constant.

NOTES AND REFERENCES

[1]H. S. D. Cole, et. al., Models of Doom (New York: Universe Books, 1973), p. 202.

[2]K. Deutsch, "Quantitative Approaches to Political Analysis: Some Past Trends and Future Prospects," in Mathematical Approaches to Politics, Alker, Deutsch, and Stoetzel, eds. (Amsterdam: Elsevier, 1973), p. 58.

[3]See, for example, W. Dray, "'Explaining What' in History," in Readings in the Philosophy of the Social Sciences, M. Brodbeck, ed. (London: Macmillan, 1968), p. 343.

[4]S. Bruchey, The Roots of American Economic Growth, 1607-1861 (New York: Harper & Row, 1965), p. 10.

[5]For an example of the difficulty in dealing with the ambiguities of the term "values," see a discussion of its varying definitions in M. Rokeach, The Nature of Human Values (New York: The Free Press, 1973), pp. 3-25.

[6]R. Rudner, Philosophy of Social Science (Englewood Cliffs, N.J.: Prentice-Hall, 1966), p. 69.

[7]M. Evans, Macroeconomic Activity: Theory, Forecasting, and Control (New York: Harper & Row, 1969), pp. 461-466.

[8]See, for example, F. Modigliani and O. Sauerlender, "Economic Expectations and Plans of Firms in Relation to Short-Term Forecasting," in Short-Term Economic Forecasting, L. Klein, ed. (Princeton: Princeton University Press, 1955), p. 261.

[9]Quoted in Bruchey, op. cit., p. 6.

[10]E. H. Phelps Brown, "The Underdevelopment of Economics," in The Economic Journal (March 1972), p. 7.

[11]J. W. Forrester, Principles of Systems (Cambridge: Wright-Allen Press, 1968), pp. 6-11 - 6-12.

[12]DYNAMO is a computer language commonly used for system dynamics models; see J. Pugh, DYNAMO II User's Manual (Cambridge: MIT Press, 1973).

[13]A full discussion of the behavior of this system can be found in N. Mass and P. Senge, "Understanding Oscillations in Simple Systems," System Dynamics Group Memo D-2045-1, August 1, 1974, MIT, Cambridge, Massachusetts.

[14]R. Dubos, So Human an Animal (New York: Scribners, 1968), p. 40.

CONSTRUCTION AND TESTING OF THE NATIONAL SOCIO-ECONOMIC MODEL: EQUATIONS FOR CORPORATE FINANCE

Alan K. Graham*
System Dynamics Group
Alfred P. Sloan School of Management
Massachusetts Institute of Technology, Cambridge, Massachusetts

The System Dynamics Group at MIT is constructing a simulation model of the US economy. The national socio-economic model addresses a wide spectrum of social and economic problems, ranging from business cycles and inflation to long-term economic growth and resource depletion. Consequently, the model structure is quite extensive. In fact, the national model is too large to be constructed and tested in a purely ad hoc fashion. This paper discusses orderly, step-by-step procedures, both in the abstract and for the specific example of equations representing corporate finance. Rather than focusing on technical details, this paper emphasizes the larger issues which motivate the choice of procedures for model construction and testing.

I. BACKGROUND

A. The National Socio-Economic Model

The national socio-economic model addresses itself to a variety of problems encompassing a variety of time horizons.** The national model contains enough short-term detail to analyze the four-year business cycle and various types of inflation.*** Over longer time-spans, the national model depicts the underlying causes of the twenty- and fifty-year cycles of economic development (the "Kuznets cycle" and the "Kondratieff cycle," respectively). Finally, the model incorporates sufficient de- to depict the national "life cycle" of economic development, growth, and eventual resource depletion.****

To organize the myriad details of so comprehensive a model, the national socio-economic model is divided into six sectors: the government, financial, household, demographic, labor, and production sectors.+ The government sector represents taxation, fiscal policy, and formulation of support programs such as social security and unemployment insurance. The financial sector represents the spectrum of financial institutions, including the Federal Reserve System. The household sector depicts the purchasing, borrowing, and savings decisions of households, as well as workforce participation. The demographic sector determines births, deaths, population age distribution, and levels of educational attainment. The labor sector represents inter-industry labor mobility. Finally, the production sector depicts corporate and other commercial activities.

The production sector describes production, ordering of capital and other factors of production, hiring and firing, price setting, borrowing, and payments. The finance equations depict the three latter activities --pricing, borrowing, and paying, as well as accounting, profitability, and the impact of financial considerations upon inventory ordering and hiring.‡ The finance equations assume a major role in several of the problems to be addressed by the national socio-economic model. Inflation by definition arises from increases in the prices of sector outputs. Financial constraints probably diminish the rate at which production can be increased during the upswings of the four-, twenty-, and fifty-year business cycles. However, the delays inherent in planning and financing new investment may in fact cause overshoots and contribute to cyclical behavior. Finally, the financial equations depict the channel through which monetary policies influence the real economy.

B. Use of Models

One must understand a model to use if effectively. The ultimate purpose of a model is to aid in designing and evaluating policies which improve the behavior of the system. To find such policies, model behavior, and the feedback loops which generate the behavior, must be well-understood. Also, a thorough understanding of model structure is necessary to translate simulation results into statements about policy alternatives in the real system: Policy analysis with a model indicates the changes which improve the behavior of the system. But

---

*Research on the National Socio-Economic Model is funded by the Rockefeller Brothers Fund. The Author is a National Science Foundation fellow.

**A more thorough discussion of the issues addressed by the national socio-economic model appears in Jay W. Forrester, "Understanding Social and Economic Change in the United States" (Proceedings of the Summer Computer Simulation Conference, San Francisco, Calif., 1975).

***Fragments of the national model have been used to investigate business cycle behavior. A brief discussion appears in Nathaniel J. Mass, "The Dynamics of Economic Fluctuations: A Framework for Analysis and Policy Design" (Proceedings of the Summer Computer Simulation Conference, San Francisco, Calif., 1975). For more details, see Nathaniel J. Mass, Economic Cycles: An Analysis of Underlying Causes, Wright-Allen Press, Cambridge, Ma., 1975.

****A preliminary study of long-term national economic development is Nathan B. Forrester, The Life Cycle of National Economic Development, Wright-Allen Press, Cambridge, Ma., 1974.

+A more extensive summary of the organization of the national model is Gilbert W. Low, "The National Socio-Economic Model--An Overview of Structure" (System Dynamics Group Memorandum D-2123).

‡A more detailed description of the finance equations appears in Nathaniel J. Mass, "The Production Sector of the National Socio-Economic Model--An Overview" (System Dynamics Group Memorandum D-2143).

effective action cannot be taken without knowledge of the real processes to which the model changes correspond;*

Unfortunately, the behavior of a model with the complexity of the national socio-economic model almost defies understanding. Completed versions of the national model will contain literally thousands of feedback loops. The task of attributing the various model behavior modes to the individual structural elements of the model is formidable. Many realtionships in the model are nonlinear. Among other problems, nonlinearity allows a single behavior mode to be caused by any one of several independent causes, depending on conditions in the rest of the model. For example, an inflationary spiral can be instigated by persistent monetary overexpansion, monopolies in key industries, rigidities in the labor market, rising expectations of standard of living, high interest (or dividend) costs, or dislocation resulting from attempts at wage and price control. Consequently, nonlinearities allow the national model to generate extremely complex behavior modes. Finally and perhaps most seriously, complex models such as the national model are in general counterintuitive. That is, the parameters which can significantly alter model behavior are not the obvious ones.**

To understand a model as complex as the national model, one can only proceed in a step-by-step manner. The national model is too complex to analyze as a single unit. In fact, many of the individual sectors, especially the production sector, appear to be too complex to analyze as a single unit. However, a fairly small group of equations within a sector does comprise a practical beginning-point for model analysis. Even a small group of equations exhibits several behavior modes and responses to exogenous inputs. If the group of equations forms more than one feedback loop, the modeler can distinguish the dominant feedback loops essential to the model behavior from the less-important loops. The modeler can also identify sensitive and insensitive parameter variations--changes in parameter values which either do or do not alter the model behavior.

After the modeler analyzes two or three small groups of equations, the groups can be combined to form a larger unit for further analysis. The interactions between the equation groups may give rise to new behavior modes, not exhibited by the individual equation groups. The connections between the equation groups form new feedback loops, which may (or may not) dominate the loops within the individual equation groups. The new feedback loops may also alter the sensitivity of model behavior to parameter changes. Based on knowledge of the smaller units, then, the modeler can analyze the new behavior modes, dominant loops, and parameter sensitivities of the larger unit in a fairly straightforward, step-by-step fashion.

C. Construction of Models

The system dynamics literature may give the impression that simulation models are initially created in their final, fully-developed form. In fact, most published models result from a fairly long and involved development. Most initial models contain "bugs." Elimination of typing errors and careless errors such as sign inversions only begins the debugging process. Simulation usually uncovers deficiencies in the formulation of the model equations. Closer scrutiny reveals that equations which initially seemed adequate in fact give responses of unrealistic magnitude or timing. Variables sometimes even move in unrealistic directions in response to unforeseen combinations of inputs. A model usually requires a significant amount of testing and reformulation before each equation gives a consistently realistic response, even under a variety of extreme conditions.

A model with the complexity of the national socio-economic model is virtually impossible and quite expensive to debug as a single unit. Such a model contains thousands of feedback loops, so a defective equation may give rise to unrealistic behavior in many different and widely separated parts of the model. Therefore, beginning with unrealistic behavior and tracing the path back to the defective equation(s) poses great difficulties. More insidiously, the national socio-economic model represents decisions and the adaptation of decision-makers to changing circumstances. The parts of the model which interact with a defective equation sometimes adaptively counteract the unrealistic responses of the defective equation. Until a very detailed examination is made, such adaptations may completely mask the symptoms of the defect.

A practical debugging method is to begin with small groups of equations. After obtaining successful compilations, test the behavior. To test the behavior, the variables defined outside a given set of equations must be initially held constant. Therefore, feedback loops which affect behavior in the real system may not operate during the initial tests. Given the possible absence of feedback loops, however, the behavior should be plausible. If not, determine the formulation changes necessary to make the behavior plausible, given the constraints of the testing procedure. When tests finally yield plausible behavior by the initial groups of equations, combine them into larger units. The new behavior modes of the larger unit may expose other equations in need of reformulation. Repeat the cycle of testing and reformulating, until the resulting behavior is plausible and still larger units of equations can be formed. In this way, construction of even large, complex models can proceed in an organized and efficient fashion.

D. Validation of Models

In the most general sense, validation is any procedure which increases confidence in the utility of a model. A variety of validity tests exist; the following six considerations apply to both model equations and model behavior.+

Purpose. Are the model structure and the model behavior consonant with the overall model purpose? Does

*An excellent example of this "translation" of model behavior appears in Chapter 7, "Interpretations" in Jay W. Forrester, Urban Dynamics, MIT Press, Cambridge, Ma., 1969.

**To be exact, the counterintuitive nature of social systems and models of social systems characterizes not the systems themselves, but our understanding of those systems. The properties of complex systems which render them counterintuitive are discussed in Chapter 6, "Notes on Complex Systems" in Urban Dynamics, op. cit.

+A more extensive discussion of model validity appears in Jay W. Forrester, Industrial Dynamics, MIT Press, Cambridge, Ma., 1961, pp. 115-129.

the model explain the causes of the problem being analyzed? Does the model contain enough structure to permit analysis of the effects of policy intervention? If not, the model has little validity as a policy-making tool.

Perceptions. Do the model structure and behavior match the perceptions of people who work within the real system? Is the model response to exogenous inputs, parameter changes, and policy changes plausible? Do the cause-and-effect relationships in the model structure correspond to perceived cause-and-effect relationships in the real system? The equations in the production sector, for example, should represent investment decisions in a form which corporate executives could recognize as being similar to the real investment decision process. The information flows in the model should represent only the information streams actually available to real decision-makers. Similarly, a corporate executive should not be able to readily distinguish the behavior of the production sector (with suitable inputs) from behavior of real corporations or sectors.

Extremes. Do the individual model equations and the whole model respond in a realistic, plausible manner to extreme conditions? Certainly, evaluating the model under conditions approximating the historical values can to some extent test the validity of the model. However, extreme conditions constitute a more difficult and hence more discriminating validity test. Suppose, for example, that the national model fails to behave plausibly under wage and price controls; the model must lack an element of real-world structure which can potentially assume major importance.

Generality. Does resetting the parameters allow the model to represent a variety of cases and historical examples? Generality is especially important in the national socio-economic model, where one set of equations depicts a variety of industrial sectors. The pricing equations, for example, should be able to represent either an auction-type market for commodities, where supply and demand determine price, or a cost-plus market, where prices are determined principally by corporate profitability, or any cases in between these two extremes.

Replication of Data. Can the model parameters be set to replicate in some sense the available statistics? The replication can be judged either by inspection of data, or by more formal statistical techniques. Inspection of data on the business cycle, for example, can indicate typical magnitudes of variables, the typical period of each cycle, and the timing of events within each cycle. A model of business cycles should replicate these features of the data.

Discovery. Do the model equations or the model behavior engender new observations about the real system? Does the model predict behavior modes which are subsequently observed in the real system?

These six validity criteria are applicable to every stage of model construction. Section II.A discusses in more detail the evaluation of individual equations with respect to purpose, perceptions, extremes, generality, replication of data, and discovery. Similarly, behavior of pieces of a model, and behavior of the whole model can be tested for validity using the same six criteria.

E. Model Construction and Testing

Model construction proceeds with alternate formulation and testing of pieces of the model, while the size of the pieces gradually increases. Validity tests are performed both on the model equations as they are formulated, and on the model behavior during the model construction. Model testing also forms the basis for understanding the model behavior, which makes possible subsequent policy design. In short, constructing, validating, and understanding a model are concurrent processes. When each portion of the model is constructed, it should also be tested for validity, and analyzed to find the underlying causes of the behavior.

The structure, validity, and policy implications of the national socio-economic model emerge quite slowly, as the step-by-step process of model-building proceeds. Models smaller and less complex than the national model can be constructed, validated, and used as policy tools as single entities, without the need for gradual construction. Smaller, less complex models possess only one or two behavior modes, and therefore can be presented and analyzed as a single entity. The national model differs sharply. The completed model will encompass thousands of feedback loops, and exhibit literally dozens of behavior modes. Each behavior mode will result from a unique set of important feedback loops. The only workable format for communicating the structure, behavior, and policy implications of such a model is to document the intermediate steps in the process of model construction. Descriptions of intermediate steps are already available for cyclic behavior in the production sectors of the national model, and for the portion of the financial sector describing monetary policy formulation.* The following section discusses the process of constructing and testing of another component of the national model, the equations which describe corporate finance.

II. AN EXAMPLE: EQUATIONS FOR CORPORATE FINANCE

Section II illustrates equation formulation, validity testing, and behavior analysis by discussing these procedures for the financing equations in the national model. For each production sector, these equations describe payments, borrowing, accounting, product pricing, and the impact of financial considerations on investment and inventory planning. The discussion does not focus on the mechanics of testing. Instead, the emphasis is on the motivations for performing each separate test, and on the use of test results in understanding the national economy. In conformance to the previous section, the discussion begins with the simplest units for analysis, the individual equations. The discussion then progresses to procedures for constructing and testing single- and double-sector models.

*Mass, Economic Cycles, op. cit., and W. W. Behrens III, "A Policy Structure for Monetary Control Decisions," PhD Thesis, MIT, 1975.

A. Equation Formulation

A relatively small group of equations can be clearly and thoroughly discussed, especially with respect to the validity testing of the equations. Indeed, the paucity of discussion on validity seems to underlie the critical reception of World Dynamics and The Limits to Growth.* A thorough and convincing presentation of model structure must necessarily contain enough information to evaluate the validity of the model structure. The general guidelines for validity testing in Section I.D provide a convenient framework for presenting the finance equations.

Purpose. How do the equations relate to the overall purpose of the model? In which behavior modes do the equations play a significant role? Is the level of detail appropriate for the model purpose?

Perceptions. What real-world phenomena does each equation represent? How are the equations consistent with descriptive, theoretic, and quantitative knowledge? The description of the equations should make sense to a person familiar with real corporate finance. The equation description should also relate the model structure to the appropriate economic theory.

Extremes. How do the equations respond to extreme conditions, and is the response plausible? For example, suppose the model equations indicate that very high interest rates cause diminution of bank borrowing and stock issue. A thorough equation description would relate this diminution to a discussion, from a managerial viewpoint, on investment planning and alternative means of obtaining funds.

Generality. Does resetting the parameter values allow the finance equations to represent the variety of characteristics of the real economic sectors? For example, the household sectors in the national model will utilize the finance equations of the production sector; the finance equations must therefore depict not only corporate financing, but also household financing. The equation descriptions should indicate how different financing characteristics, such as risk factors and borrowing habits, are reflected by parameter variations.

Replication of Data. Each equation defines an output variable in terms of one or more input variables. Cross-sectional or time-series data is often available for both input and output variables. Each equation should in some sense replicate that data. Replication is fairly straightforward to test for equations with one input variable: For known values of the input variable, does the equation give an appropriate value of the output variable? Testing replication of data for equations with more than one input variable, however, requires considerable care, due to potential statistical problems.**

Discovery. Do the equations specify relationships which are not usually considered? The formulation of new relationships frequently arises from consideration of extreme conditions. For example, Urban Dynamics posits a relationship which restricts migration into a city when the city's housing stock is already overcrowded.† Historically, this relationship has not been critical to urban behavior. Yet this relationship is central to the policy recommendations.‡ A realistic model should not depict only those relationships which have been important in the past. A realistic model should also depict relationships which potentially assume major importance under conditions substantially different from the past.

B. Single-Sector Testing

Two goals motivate the testing of the finance equations within a single production sector: validation and understanding. Showing that the model behavior is plausible and realistic, even under extreme conditions, enhances the validity of the model. Understanding the role of the finance equations in the dynamics of the single sector establishes a groundwork for analyzing the interactions between the real and money sectors of the economy. The discussion below identifies several specific instances of the effect of financial variables upon productive activity.

Behavior testing begins with tests on the finance equations in isolation. Writing the equations in a completely separate model, however, requires too much effort. A more efficient method of testing involves writing the equations for the entire production sector, but setting parameters to immobilize selected parts of the sector. Setting parameters to maintain inventories and production at constant values in effect allows testing the finance equations in isolation.

---

*Jay W. Forrester, World Dynamics, Wright-Allen Press, Cambridge, Ma., 1971, and Dennis L. Meadows et al., The Limits to Growth, Potomac Associates, Washington, D.C., 1972. A discussion of the misconceptions engendered by these relatively short presentations appears in Jay W. Forrester, Nathaniel J. Mass, and Gilbert W. Low, "The Debate on World Dynamics: A Response to Nordhaus," Policy Sciences 5: pp. 169-190, June, 1974. In contrast to the shorter presentations, the much more thorough report on the same subject matter has been received virtually without comment (Dennis L. Meadows et al., Dynamics of Growth in a Finite World, Wright-Allen Press, Cambridge, Ma., 1974).

**Within a dynamic feedback system, variables are commonly highly correlated. Causality is accordingly difficult to infer in an equation with multiple inputs: If several input variables vary at once, changes in the output cannot be easily attributed to any single variable. Forrester, Low, and Mass (op. cit., pp. 171-177) give an example in which the correlation between an input and an output variable indicates the opposite of the true structural relationship. The statistical problems can seemingly be overcome by formal estimation procedures. However, these procedures are also vulnerable to considerable error; see Peter M. Senge, "An Experimental Evaluation of Generalized Least Squares Estimation," System Dynamics Group Working Paper D-1944-6.

†For more extensive discussion, see Alexander G. Makowski, "Housing and Migration in Urban Dynamics," in Walter W. Schroeder III, Robert E. Sweeney, and Louis E. Alfeld, eds., Readings in Urban Dynamics, vol. II, Wright-Allen Press, Cambridge, Ma., 1975, Reading 4.

‡Urban Dynamics, op. cit., Chapter 5, "Urban Revival."

Validity testing focuses on examining the plausibility of model behavior, even under extreme conditions. Even when the production rate and inventories are held constant, the financing equations can respond to inputs by varying payment schedules, borrowing, wages and prices, and ultimately defaults. Thus, the single-sector model should yield moderately plausible responses, even to extreme values of interest rates, loan availability, wage demands, or cost increases.

Model testing should also be used to enhance the understanding of the model's dynamics. Tests can indicate which loops in the model are important to its short-run and long-run behavior. Knowledge of the model's dominant loops in turn reveals potentially sensitive parameters. Most importantly, however, model tests evaluate behavior modes of the individual unit being tested which are important to the dynamics of the entire model. For example, a variety of theories exist which explain how and why a recession becomes a depression. One hypothesis is that depressions result when a downturn progresses to the point where business failures and the resulting defaults put a significant strain on the economy as a whole. A critical feature of this hypothesis is the existence of a "chain reaction" of business failures, where financial difficulties in one sector cause it to default on accounts payable, which in turn engenders financial difficulties in other sectors.

One important set of tests, therefore is to determine the parameter values and internal financial conditions for which defaults on accounts receivable can cause comparable defaults on accounts payable. In other words, the tests define the circumstances in which business failues can propagate throughout the economy in a "chain reaction." If the tests reveal that the "chain reaction" can occur only in absurd circumstances, then the investigator must look elsewhere for an explanation of depressions. But if the tests indicate that a "chain reaction" can occur under plausible circumstances, the tests provide both a basis for further investigation, and a preliminary set of economic indicators which, at least on the model, indicate vulnerability to "chain reactions" of business failures.

The next stage in single-sector behavior testing allows the production rate and the ordering and inventories of factors of production to vary, but maintains the fixed endogenous inputs such as orders for finished products, interest rates, and delivery delays. Before going on to evaluate new dynamic hypotheses, the results of the previous testing should be tested in the new model (or more precisely, in the same model with parameters which activate previously immobilized equations). New feedback loops are activated which may alter loop dominance and parameter sensitivity. For example, a tight cash position reduces inventory ordering and accounts payable, which can somewhat alleviate the tight cash position. This feedback loop, by further cushioning the sector's cash balances, may reduce the importance of other loops which control cash balances. Also, the new feedback loop may alter the circumstances in which a "chain reaction" of business failures can occur.

Freeing the production rate allows the sector to experience four-year and twenty-year business cycles. From the viewpoint of monetary policy formulation, it is important to know the extent to which financial variables influence business cycles. Monetary policy does influence the production sector. However, the sector attempts to counteract the influence of restrictive monetary policies by delaying payments of accounts payable, borrowing despite high costs, raising prices, or cutting costs. Suppose sudden changes in interest rates cannot initiate business cycles due to cushioning by the negative feedback loops in the financial sector. Then a rapidly-changing monetary policy probably cannot effectively counteract business cycles, either. The response of the single sector to interest changes therefore assumes major importance to policy-makers.

The effect of financial variables on business cycles is of interest in economic theory. Prewar and postwar business cycles exhibit different turning points, damping, and period; many of the differences may be due to the different interest rates and monetary conditions before and after the way. Since postwar interest and dividend payments in general comprise a larger share of a firm's total revenues, rapid expansion should be more difficult to finance than during the prewar period. This lessened ability to expand might account for the smaller amplitude of the average postwar business cycle. This hypothesis is easily tested by examining a single sector's cyclical behavior under constant easy money conditions and under constant tight money conditions.

Single-sector testing also forms a basis for understanding inflation. While inflation obviously does not arise exclusively from corporate actions, the response of sector prices to various sector inputs obviously forms a critical link in the overall structure responsible for inflation. For example, consider the sector's response to factor cost increases. To the extent that either other factors can be substituted for a costly factor, or costs can be temporarily met through borrowing, increasing factor costs will not be passed on to prices. If the substitution takes several years and borrowing is impractical, however, prices must increase in the interim to cover costs.

The effectiveness of anti-inflationary policies depends critically upon the promptness with which each production sector passes on cost increases as price increases. If the production sectors pass on cost increases only slowly, the wage- (or cost-) price spiral cannot persist independent of "outside" causes of inflation such as oil price increases and government deficit spending. If so, inflation can be controlled by controlling the "outside" causes of inflation. If the production sectors pass on cost increases quickly, the wage-price spiral can persist for some time, even in the absence of "outside" price stimuli. In this case, inflation can only be controlled by some means which restrains price rises in each sector.*

Single-sector testing can indicate the extent to which inflation arises from economic change or dislocations. In a corporate setting, prices seem more readily

*In the case of quick cost pass-through and persistent wage-price spirals, wage-price controls must be completely effective to be effective at all: any price rise generates pressure for price rises in other sectors. Unless controls are completely effective, some prices will rise, generating still more pressure to raise prices or wages.

raised than lowered. Therefore, as either demand or costs fluctuate, price rises are not matched by price declines. Of course, prices cannot rise very far above costs, because the high profit margin would eventually encourage new firms to enter the market at lower prices.*

Current macroeconomic analysis seems to focus on the use of aggregate fiscal and monetary policy instruments. Yet these instruments may not be appropriate for controlling inflation, if inflation is in large measure due to sector-by-sector fluctuations in demand or cost. If the current inflation, for example, is due to a capacity constraint in the energy sector, aggregate policy instruments can do little to solve the underlying problem. In fact, using monetary controls to combat inflation has probably not only significantly diminished the cost of energy, but has inflicted considerable damage on the construction industry as well. The balance between aggregate economic policies and policies for individual sectors is important; single-sector testing can give some indications on appropriate balance.

The final stage in single-sector testing allows the previously exogenous inputs to respond to conditions within the sector. Orders for the sector's output could respond to price and delivery delay. Interest costs could respond to the level of corporate borrowing and the risk of loan defaults. As before, several new feedback loops are formed which could modify the results of earlier testing. The earlier results need to be checked with the additional structure. The additional structure also makes possible new behavior modes.

The Limits to Growth research (op. cit.) indicates that industrial output and population may decline within the next century, either unintentionally or intentionally. This decline poses problems not frequently encountered in the present mode of long-term economic growth: What happens when a sector's output and capital plant must shrink? The diminution need not be gradual; in fact, quite dramatic changes are possible. For example, consider the financial standing of a sector for whose output the demand is shrinking. In the past, if sector operations were profitable, the sector could obtain the financing it needed. However, as demand shrinks, fixed costs reduce profit margins and weaken the sector's financial position. Banks will usually still lend funds to such a sector, however, in order to allow the company or sector to acquire capital goods needed to operate effectively, and to repay both new and old loans. Past some point, however, banks will not extend further credit to a risky sector, or will extend credit at high cost. The high cost or lack of adequate credit further weakens the sector. Once this "financial watershed" is passed, the financial condition of the sector deteriorates rapidly. Single-sector testing can indicate the situations in which this deterioration occurs, and possibly how to avoid it.

C. Double-Sector Testing

At present, the dynamics of even the production and inventory portions of a two-sector model are not well-understood. Yet to be explored are the various configurations of two sectors purchasing output from themselves or each other, and phenomena such as entrainment. Entrainment occurs when cycles in one sector give rise to comparable, in-phase cycles in another sector. Cycles in the individual sectors of the economy must somehow be entrained, if the aggregate national economy can experience uniform business cycles. So until the dynamics of production and inventory cycles in the two-sector model are better understood, assessment of the impact of financial considerations on cyclic behavior is quite difficult.

As in the one-sector case, the analysis of the two-sector model with respect to the financial equations begins by making constant the inventories and production rates in each sector. The partial two-sector model can then be used to investigate tentative conclusions of the earlier tests. For example, the one-sector tests give the circumstances in which a sector transmits almost all cost increases to increased prices. Two-sector testing expands upon the analysis by determining the circumstances (parameters and input values) in which two sectors can generate a true cost-price spiral--prices of the first sector increasing the costs of the second, and vice versa. Another example of expanding upon the one-sector results is the transmission of financial difficulties. A two-sector model could show the circumstances in which financial difficulties in one sector cause financial difficulties in the other sector via defaults or delays on accounts payable, which further aggravate the difficulties in the first sector.

A further stage of two-sector testing involves two fully active sectors, including variable production, factor ordering, and product ordering. Then couple the two sectors to one another by allowing interest rates and loan availability to respond to the sum of the loan demands from the two sectors. This coupling allows tests of entrainment through interest rates and loan availability.

D. Conclusion

At this point, an important caveat must be stated. The preceding text has emphasized the advantages of constructing and testing a model one piece at a time. This step-by-step procedure is efficient only when the purpose and overall structure of the model are already well-defined. Without some idea of the problems to be analyzed and the policies to be tested, there is no basis for either terminating the modeling process, or choosing pieces to add to a model. Adding pieces to a model in the vague hope of making the behavior more realistic is a never-ending process.

Constructing, validating, and understanding models one piece at a time is very practical for models whose purpose and overall structure are well-defined. Construction is more easily performed with well-understood pieces. Model behavior is more easily understood when the behavior results from a simple model. A step-by-step approach permits the national model and understanding of the behavior of the national model to grow gradually, without the necessity of dealing with the full complexity of a completed model.

In summary, each portion of the national socio-economic model should be regarded likewise as a model, with its own strengths and weaknesses. Each model exhibits its own characteristic behavior, and yields its own implications for policy design. The step-by-step procedures described above consist of nothing more than learning as much as possible from each of these models.

*The single aggregate production sector represents competition in markets by the values of the parameters governing pricing and sector expansion. The production sector depicts entry of new firms into a market during periods of high profitability by increases in orders for factors of production, and constraints on further price increases.