IIASA REPORTS

ISSN 0250-7625

International Institute for Applied Systems Analysis

Volume 3 Number 2 April–June 1981
Completing Volume 3

IIASA
REPORTS

Volume 3 (1981)
Completing Volume 3

Cover

Photograph: Robert A. Duis

International Standard Serial Number 0250-7625

IIASA Reports presents research conducted at IIASA that has been independently reviewed before publication. However, the views and opinions it expresses are not necessarily those of the Institute or the National Member Organizations that support it or of the institution that sponsored the research.

ENERGY IN A FINITE WORLD: IIASA's Study of Global Energy Supply and Demand through 2030

"It could be done." This is the good news from a major IIASA study of the prospects for meeting the global demands for energy over the next fifty years. During this period the world's population will double to eight billion people, and, even with only modest economic growth and extensive conservation, the global energy demand is likely to expand to three or four times today's level. Nevertheless, IIASA's multidisciplinary study team of scientists from 20 countries, both East and West, has concluded that the technology and resources can be available to satisfy this increased demand.

The study's not-so-good news is that, in order to meet this growing demand, the world must make full use of all available energy resources: coal, oil and gas, solar, renewables, and nuclear. Dirtier and more expensive fossil resources and vast quantities of synfuels will have to be developed, as well as large-scale solar plants and nuclear breeder reactors. Small-scale solar installations and renewable resources must play a growing role, too, but can only satisfy a modest fraction of the total demand during the next half century.

IIASA's Energy Systems Program Group has reported these findings, and the detailed analysis supporting them, in *Energy in a Finite World: A Global Systems Analysis* (Ballinger Publishing Company, Cambridge, Massachusetts, 1981, 837 pages); *Energy in a Finite World: Paths to a Sustainable Future* (same publisher, 225 pages) presents a shorter account for the general reader; an *Executive Summary* (74 pages) is available from IIASA on request. The IIASA analysis is the first comprehensive global long-term examination of the energy future, and the first in which scientists from East and West have collaborated.

By using a consistent model of worldwide energy supply and demand, it avoids the common tendency of separate national studies to assume that sufficient imports will always be available, without comparing the demands of all countries against the likely supplies. By looking fifty years ahead it accounts for the time it takes the energy system to undergo fundamental changes.

The principal goal of the study was to identify strategies for the transition from a globe reliant on oil and gas to one served by sustainable sources of energy. But the original expectation that this could be accomplished within a 50-year horizon turned out to be too optimistic. Instead, the IIASA group found that there will have to be two transitions. The first, from relatively cheap and clean conventional sources of oil and gas to more expensive and dirtier unconventional ones will continue through 2030. The second, to the essentially infinite supplies of solar, nuclear, and renewable energy, will not be completed until late in the next century. But such a system would be sufficient to sustain the then anticipated global population of about 10 billion persons for many centuries.

While reaching conclusions more reassuring than some previous global studies, the analysts are not completely confident about the chances that the promising paths will be followed: "The transition from the present fossil era to an era based on inexhaustible energy resources will not be straightforward. We cannot even be sure it is possible. At the very least, it will require that national energy policies, corporate energy policies, and personal energy behavior be conceived with as clear an understanding of their relationship to the global energy problem as possible. For better or worse, we cannot isolate ourselves."

They point out that all future energy paths have their costs: lower energy use threatens more severe economic difficulties, higher energy use permits greater economic development, but poses more severe environmental dangers.

Liquid fuel supply is the "energy problem within the energy problem". Even though oil supplies will increase through exploiting costlier and dirtier resources, such as oil shales and tar sands, they will be insufficient to match the rapidly expanding, and irreducible worldwide demand for liquid fuels for transportation. Vast quantities of coal will have to be liquefied. With 90 percent of the world's coal supplies in the USA, USSR, and China, these nations will play a central role in the world market that will be needed to match supplies with demand. A similar market may also develop for the synthetic fuels produced from oil shales and tar sands, located primarily in the Americas and China.

Even so, the analysts anticipate that, in the first decades of the next century, the Persian Gulf will still be supplying large quantities of oil to the world. However, its principal customers will lie in Western Europe and Japan and in the developing nations of Africa and Southeast Asia. The Americas, Eastern Europe, and China will not be net importers of oil; they will be able to satisfy their liquid fuel demands with their own oil, gas, and coal resources.

However, the authors warn that the increased use of fossil resources could be constrained by the resulting carbon dioxide releases to the atmosphere, which some scientists believe will lead to climatic changes.

The transition from today's oil, gas, and coal to fossil resources requiring substantial transformation before use and the development of renewable resources entails tremendous capital investments. While the industrial world is expected to be able to cope with this huge capital demand, the developing countries may find it difficult to provide the necessary funds.

The most precious – and scarce – resource, however, is time. In the past, new primary energy sources, such as coal, oil, and gas, have required some 100 years to increase their global market share from one to fifty percent. Therefore, the main point in solving the energy problem is not which energy resources should be chosen, but how fast we will be able to develop them. For example, large-scale solar energy deployment, such as solar power plants in desert areas, has not yet reached sufficient technological maturity to make a major impact on the global scale within the next fifty years. Solar power is expected to reach its full potential only in the second half of the next century.

Energy in a Finite World provides the factual basis for designing a world energy strategy to reach the goal of a global sustainable energy system. By identifying the problem areas, it can help politicians and policy makers reach decisions that will provide for an orderly growth of energy resources to satisfy growing world needs in peace.

The study investigates the global energy problem on three different levels:

- First, it explores the maximum global potential of the various global energy sources: oil and gas, coal, nuclear, solar, and renewables.
- Second, it investigates two scenarios – one with a high and one with a low

energy demand. In addition to these two bench-mark scenarios, three supplementary cases look at alternative paths of development: stronger deployment of nuclear power; a nuclear moratorium; and a very-low-demand development based on an unchanged average per capita energy consumption over the next fifty years (approximately two kilowatts per capita).

- Third, it identifies a number of conclusions that are relevant to globally oriented policies toward a sustainable future.

Energy in a Finite World does not provide easy answers, but for the first time it gives a global framework for decision makers all over the world. As Professor Häfele puts it: "It could be done, but only with pain and at high cost. However, if we fail to meet the challenge of the energy squeeze within the next couple of years, we may have to pay a much higher price in the long run. Time is our most scarce and valuable resource."

This study was supported primarily by funds from IIASA's National Member Organizations. Significant additional support, however, came from the United Nations Environment Program, the Volkswagen Foundation in the Federal Republic of Germany (FRG), the FRG Ministry of Research and Technology, and the Austrian National Bank. Major parts of the study were carried out in close cooperation with scientific institutions throughout the world, including, for example, the Meteorological Office (Bracknell, UK), the Nuclear Research Center (Karlsruhe, FRG), the National Center for Atmospheric Research (Boulder, Colorado, USA), the Siberian Power Institute (Irkutsk, USSR), the International Atomic Energy Agency (Vienna, Austria), and the Institute of Energy Economics and Law (Grenoble, France).

For ordering information please see inside back cover

ESTIMATION OF FARM SUPPLY RESPONSE AND ACREAGE ALLOCATION: A Case Study of Indian Agriculture

N.S.S. Narayana and Kirit S. Parikh
International Institute for Applied Systems Analysis, Austria

SUMMARY

Some of the most important decisions in agricultural production, such as what crops to grow and on how much land to grow them, must be made without certain knowledge of future rainfall, yields, and prices. In this report we model the land allocation decisions of Indian farmers as a significant first step in developing a model for Indian agricultural policy. The approach that we have adopted is consistent with the premise that farmers behave rationally and react to circumstances in a way that maximizes their utility in the context of opportunities, uncertainties, and risks as perceived by them.

After a brief review of the approaches available for estimating farm supply response, we summarize a few relevant studies, which are constructed largely after the traditional Nerlovian model, based on adaptive expectations and adjustment schemes. Significantly, however, the model seems to involve a serious error of specification with respect to the formulation of the price expectation function. Nerlovian specification does not separate past, actually realized prices into "stationary" (expected) and random components, and it attaches the same weights to the two components for predicting expected prices.

The model described in this report deviates from the traditional Nerlovian model in two principal respects:

- *We estimated acreage response for different crops by using expected revenue instead of expected prices as a proxy for expected profits.*
- *We formulated an appropriate revenue (or price, as the case may be) expectation function for each crop by clearly identifying the "stationary" and random components involved in past values of the variable and by attaching suitable weights to these components for prediction purposes. We postulated an auto-regressive integrated moving average (ARIMA) model for this purpose and used Box--Jenkins methodology in estimating these functions.*

In our study we considered nearly all crops grown in India. On the basis of sowing and harvesting periods in different states, we drew up an overall substitution pattern among

crops at the national level. This pattern permitted us to classify the crops into ten groups; the crops in different groups are usually grown in different soils, seasons, or both. The essential data for estimating the acreage response consist of area, production, yield, irrigation, prices, and rainfall.

We then inserted into the Nerlovian model the estimated revenue expectation functions for different crops and estimated the acreage response equations. Later we formulated an area allocation scheme so that the individually estimated areas of different crops would add up to the exogenously specified total gross cropped area in the country. Finally, we subjected all of the estimated equations to a validation exercise to judge the model's performance, particularly its ability to predict turning points.

---

1 THE PROBLEM AND ITS IMPORTANCE

Any analysis of agricultural policy needs to deal with the problem of affecting the supply of agricultural outputs. For policy purposes, not only the levels, but also the composition, of outputs are relevant. Agricultural supply, however, is the result of the decisions of a large number of farmers. How do farmers decide what and how much to produce? What policy instruments and other factors affect their decisions? We must understand these questions if we hope to devise a successful policy.

An important characteristic of agricultural production is the time lag that it involves: outputs are obtained months after planting operations are begun. After planting has been completed, farmers have comparatively little control over output.

The most important decisions -- what crops to grow and on how much land – must be made without certain knowledge of future rainfall or harvest prices. How do farmers form their expectations about these factors? How do their expectations affect their crucial decisions about land allocation?

In this report we investigate these issues in India. Modeling the land allocation decisions of Indian farmers is an important first step in developing a model for Indian agricultural policy. K.S. Parikh (1977) has described the framework of the full model, which is a computable, general equilibrium model.

We start with the premise that farmers behave rationally and that rational farmers should react in a way that maximizes their utility within the context of the opportunities, uncertainties, and risks that they perceive. Our approach is consistent with this premise. We have estimated our model econometrically, using Indian data covering the period from 1950 to 1974. The model states that farmers' desired allocation of their land among competing crops depends on rainfall and on the relative revenue that they expect to derive from different crops. Moreover, various constraints may restrict the rate at which the farmers can adapt to a desired new cropping pattern.

We have used expected revenue rather than expected prices, not only because expected revenue is theoretically more satisfactory (farmers must observe that in good years prices fall), but also because a great deal of uncertainty is associated with yields. Expected revenue is used as a proxy for expected profits because adequate data for crop-specific costs and profits are not available, and for farmers who operate with a fixed amount of total available inputs (an amount that is less than the profit-maximizing input level), maximizing profits and maximizing revenue give nearly the same results.

The model may be used as part of a year-by-year, simulation-type, price-endogenous, computable, general equilibrium model. We have carried out validation exercises to test its performance in simulating the area allocation system developed.

In the next section we discuss certain methodological issues. A review of literature follows in Section 3. In Section 4 we describe our experience with the estimation of the Nerlovian model on acreage responses, the estimation of crop revenue expectation functions based on the Box–Jenkins methodology, and the modified acreage response model. In Section 5 we describe the validation exercises. A discussion of policy implications and conclusions follows in Sections 6 and 7.

2 POSSIBLE APPROACHES TO MODELING SUPPLY RESPONSE

We have followed a two-stage approach to modeling supply response. In the first stage, which is described in this report, farmers allocate their land to different crops. This is followed by a second stage in which, given the areas, yields are determined. The first-stage model is econometric. The second-stage model may be a programming one in which farmers allocate the inputs and factors other than land to different crops in order to maximize profits. Alternately, yields in the second stage may be estimated econometrically as a function of inputs and rainfall.

Why have we followed a two-stage procedure instead of one in which all allocation decisions (of land, as well as of other factors and inputs) are made simultaneously? In a one-stage procedure, two broad approaches are possible. One is to develop a programming model in which area allocation is internal; the other is to have an econometric estimate of the output levels themselves as supply functions.

Each alternative has limitations. A programming approach leads to a corner solution, in which land is allocated to one crop, unless the area allocations are constrained either explicitly or through production functions in which there are diminishing returns to area devoted to one crop. A corner solution may also be avoided by introducing measures of uncertainty regarding the output of various crops. It is sometimes suggested that explicit constraints on areas prescribed exogenously are acceptable or even desirable, particularly when farmers consume a large amount of their output themselves. This argument, however, implicitly assumes either that farmers' allocation decisions are so complex that they cannot be modeled or that farmers have so little choice in allocating land to different crops that the arbitrariness of explicit area constraints is tolerable. These assumptions are questionable and need to be tested empirically, for even farmers growing food largely for self-consumption should not be insensitive to changing prices and profitabilities. In self-consumption, where the farmer essentially sells to and buys from himself, the trade margin on that amount accrues to the farmer himself. Taking this into account, a rational farmer should want to maximize expected profits, including margin on trade for self-consumption. Similarly, the perverse relationship of marketable surplus to prices (marketable surplus going down as prices rise; see Krishnan 1965) can also be consistent with conventional economic theory. As higher prices for his products make him richer, the farmer might want to consume more of his own product. These arguments suggest that one should consider modeling farmers' land allocation decisions before one adopts arbitrary constraints.

An alternative method of avoiding corner solutions in a programming model is to introduce diminishing returns to size of area devoted to a crop. Empirical estimates of such production functions are not easy to make and are not generally available. Moreover, the data required to make such estimates are not plentiful. This is therefore a hard procedure to follow. The difficulty of introducing in a programming model uncertainties regarding various crops is essentially that of identifying separately the variations in yield levels resulting from input levels and weather.

Estimating an econometric output supply function is unsatisfactory for a policy simulation model because only the final outcome of a number of decisions is estimated. The estimation thus provides less flexibility in changing certain parameters in the model. For example, the impact of new high-yield varieties might be hard to assess in such a framework. We have therefore followed a two-stage model.

3 A BRIEF REVIEW OF LITERATURE ON SUPPLY RESPONSE

Most empirical research on estimating farmers' acreage response is based on direct application, minor modification, or extension of the celebrated work of Nerlove (1958). Nerlove distinguishes three types of output changes: "(1) in response to changes in current prices which do not affect the level of expected future prices, (2) in immediate response to a change in the level of expected future prices, and (3) in response to a change in the expected and actual level of prices after sufficient time has elapsed to make full adjustment possible."

Of these, output changes of the first type may be limited for two reasons. First, a sudden change in output based on sudden changes in input–output prices may be difficult to achieve. Second, if the change (increase or decrease) is only a short-term phenomenon, such quick and frequent output changes may be quite costly. Hence we ignore output changes of the first type and are left with the three essential ideas of the Nerlovian model: (1) over time, farmers keep adjusting their output toward a desired (or equilibrium) level of output in the long run, based on expected future prices; (2) current prices affect output only to the extent that they alter expected future prices; and (3) short-term adjustments in output, which are made keeping the long-term desired level of output in mind, may not fully reach the long-term desired level because constraints on the speed of acreage adjustment may exist.

Nerlove's model is as follows:

$$X_t^* = a_0 + a_1 P_t^* + a_2 Z_t + U_t \tag{1}$$

$$P_t^* = \beta P_{t-1} + (1 - \beta) P_{t-1}^* \qquad 0 < \beta \leqslant 1 \tag{2}$$

$$X_t = (1 - \gamma) X_{t-1} + \gamma X_t^* \qquad 0 < \gamma \leqslant 1 \tag{3}$$

where

X_t^* is the long-term desired (equilibrium) acreage of the crop in period t

X_t is the actual acreage

P_t^* is the expected "normal" price

P_t is the actual price

Z_t is any other relevant variable (say, rainfall)

U_t is a random residual

β is the price expectation coefficient

γ is the acreage adjustment coefficient

Given that $0 < \beta \leq 1$, eq. (2) implies that the current expected price P_t^* falls somewhere between the previous year's actual price P_{t-1} and the previous year's expected price P_{t-1}^*. That is, the current year's expected price is revised in proportion to the difference between actual and expected prices in the previous year. If $\beta = 0$, the expectation pattern is independent of the actual prices, and only one expected price for all time periods exists. If $\beta = 1$, the current year's expected price is always equal to the previous year's actual price.

The restriction $0 < \beta \leq 1$ is an essential one. The value of β indicates the nature of the movement of price expectations over time as actual prices are observed. If $\beta < 0$ or $\beta > 2$, the price expectation pattern represents a movement away from the actual price movement. Moreover, when $\beta > 1$, the weight for P_{t-1}^* becomes negative, which does not seem aesthetically appealing. Some researchers, such as Cummings (1975), have presented empirical results that do not satisfy the condition $0 < \beta \leq 1$.

Equation (3) also implies a similar process of acreage adjustment. Farmers adjust their acreage in proportion to the difference between the desired or long-term equilibrium level and the actual acreage level during the previous period. Again, a meaningful interpretation requires that $0 < \gamma \leq 1$, for $\gamma < 0$ implies that a farmer allocates less area in time t than that in time $t - 1$, while in fact he desires to have more area (assuming that $X_t^* > X_{t-1}$) and $\gamma > 1$ implies overadjustment.

Equations (1), (2), and (3) contain the long-term equilibrium and expected variables that are not observable. However, for estimation purposes, a reduced form containing only observable variables may be written (after some algebraic manipulation) as follows:

$$X_t = a_0\beta\gamma + a_1\beta\gamma P_{t-1} + (1 - \beta + 1 - \gamma)X_{t-1} - (1 - \beta)(1 - \gamma)X_{t-2}$$
$$+ a_2\gamma Z_t - a_2(1 - \beta)\gamma Z_{t-1} + \gamma[U_t - (1 - \beta)U_{t-1}] \quad (4)$$

Underlying the reduced form (eq. (4)) are the hypotheses and assumptions described above, although it might be possible to arrive at the same reduced form under a different set of hypotheses and assumptions. Unless the structural parameters are identified and found satisfactory, a good fit for the reduced form is hard to interpret.

Fisher and Temin (1970) give an example of a reduced-form equation (notation changed and trend variable t added here) obtainable by different sets of hypotheses:

$$X_t = a_1 + a_2P_{t-1} + a_3t + a_4X_{t-1} + U_t \quad (5)$$

They say that one may arrive at eq. (5) in at least three different ways. First, eq. (5) can be modified and rewritten to express X_t as a function of past prices, which then means that current acreage is related to past observed prices. Second, farmers may conceive of a desired level of acreage – say, X_t^* – knowing P_{t-1}, but may somehow be unable to achieve that level. If

$$X_t^* = a_1^* + a_2^*P_{t-1} + a_3^*t + U_t^*$$

and

$$X_t - X_{t-1} = \mu(X_t^* - X_t) + W_t \qquad 0 < \mu \leq 1$$

it is possible to arrive at eq. (5) after substitution. Third, whatever their adjustment ability may be, farmers may make decisions on the basis of the price that they expect from their observations of actual prices. If

$$X_t = a^* + a_2^* P_{t-1}^* + a_3^* t + V_t$$

and

$$P_t^* - P_{t-1}^* = \mu(P_t - P_{t-1}^*) \qquad 0 < \mu \leqslant 1$$

then again from these two relations X_t can be expressed as a function of past prices.

In the previously mentioned cases, these hypotheses lead to reduced forms that are not distinguishable by observation. The Nerlovian case corresponds to a situation where the last two hypotheses were made together.

Equation (4) involves some estimation problems that we should mention briefly here. Suppposing that there is no Z_t variable in eq. (1), the reduced form becomes

$$X_t = a_0\beta\gamma + a_1\beta\gamma P_{t-1} + (1 - \beta + 1 - \gamma)X_{t-1} - (1 - \beta)(1 - \gamma)X_{t-2}$$

$$+ \gamma[U_t - (1 - \beta)U_{t-1}] \tag{6}$$

Then $\beta\gamma$ (i.e., the product of β and γ), but not β and γ separately, can be obtained from the quadratic equation formed from the coefficients of X_{t-1} and X_{t-2} of eq. (6). Using the estimate of $\beta\gamma$, however, an estimate of a_1 clearly can be obtained. Hence, even though the adjustment and expectation parameters β and γ are not identified separately, the long-term elasticity with respect to expected price may still be known.

This difficulty of parameter identification cannot be overcome, even by introducing another variable Z_t into the system. As can be seen from eq. (4), such an introduction yields separate, but not unique, estimates of β and γ. However, by postulating a suitable expectation pattern, one might be able to solve this difficulty. In the Nerlovian system, farmers have expectations only about the price variable. Actually, farmers might have simultaneous expectations about such other variables as yield or rainfall. Their area allocation decisions would follow from these expectations.

During the last decade and a half, Nerlove's model has inspired a great deal of empirical research (see Askari and Cummings 1976) in a number of countries, including India, with respect to estimating the acreage response of farmers to price movements. A review of relevant literature, including modifications and extensions of the Nerlovian model and occasional comments about the estimation problems involved, follows.

R. Krishna (1963) made one of the earliest attempts to apply a Nerlovian approach to Indian data. His model, simply an area adjustment supply model, includes irrigation, rainfall, relative price, and yield variables. He does not distinguish between actual and expected prices, which implies that farmers have full knowledge of what prices are going to be.†

†Behrman (1968) gives a critical analysis of this model.

Narain's study (1965) on the impact of price movements on areas under selected Indian crops is not based on a Nerlovian approach but on graphical analysis. As it is not based on econometric analysis, the usual estimation problems disappear in Narain's work, but comparison of his approach and results with those of other researchers is difficult.†

Cummings (1975) writes the reduced form (eq. (4)) in the following way:

$$A_t - (1-\beta)A_{t-1} = a_0\beta\gamma + a_1\beta\gamma P_{t-1} + (1-\gamma)[A_{t-1} - (1-\beta)A_{t-2}]$$
$$+ a_2\gamma[Z_t - (1-\beta)Z_{t-1}] + \gamma[U_t - (1-\beta)U_{t-1}] \tag{7}$$

He estimates eq. (7) for a range of specified values of β and selects that value of β "for which the regression error sum of squares is minimized." Two points should be noted. First, according to Cummings, the price expectation coefficient "can be reasonably assumed to fall within the range of zero to two." No justification is provided for assuming β to be greater than one. Second, to take care of autocorrelation, Cummings employs the Cochrane–Orcutt technique, which uses a first-order autocorrelation scheme on the disturbance terms.

If eq. (7) is estimated, it means that the following is assumed to be true:

$$U_t - (1-\beta)U_{t-1} = \rho[U_{t-1} - (1-\beta)U_{t-2}] + V_t \tag{8}$$

With the usual assumptions for V_t and ρ, eq. (8) implies a second-order scheme of autodisturbance for U_t, which is the basic disturbance term in eq. (1). Cummings explains neither the second-order autocorrelation scheme of U_t nor the first-order one shown in eq. (8).

Madhavan (1972) pays explicit attention to deriving eq. (1), the first equation of the Nerlovian scheme. He formulates a Lagrangian to maximize farmers' net income:

$$J = \sum_i P_i Y_i - \mu H(Y_1, \ldots, Y_m)$$

where Y_i is the production function for the ith crop and H is the same for the farm as a whole. Setting the partial derivatives to be zero and imposing the marginality conditions

$$\frac{(\partial Y_i/\partial X_i^*)}{(\partial Y_j/\partial X_j^*)} = \frac{P_j^*}{P_i^*} \tag{9}$$

he derives

$$\log X_i^* = a_0 + a_1 \log (P_j^*/P_i^*) + a_2 \log Y_i^* + a_3 \log Y_j^* + a_4 \log X_j^* + U_i \tag{10}$$

where X_i^* is the desired acreage of the ith crop, X_j^* is the desired acreage of the jth crop, and P^* and Y^* are the expected levels of prices and yields. This formulation is interesting because it is a consequence of the maximization procedure. Madhavan also introduces

†Lipton (1966) makes further comments on this study.

competing crops and relative yields. With respect to expectations, however, he assumes current expectations to be the previous year's actual values.

The next step in this field of research was to incorporate the elements of risk and uncertainty. In a case study of four major annual crops in Thailand from 1937 to 1963, Behrman (1968) attempts to capture the influences of variability of prices and yields on supply response functions. Along with such variables as population and the death rate from malaria, he introduces the standard deviations of price and yield in the three previous periods to give an idea of farmers' reactions to risks. However, Nowshirvani (1971) points out that Behrman's analysis was an empirical exercise without an explicit theoretical model. He also contends that Behrman's procedure is somewhat unsatisfactory because "the Nerlovian price expectation model is inconsistent with a changing variance of the subjective probability distribution of prices."

Nowshirvani develops a theoretical model for farmers' decisions on land allocation that accounts for uncertainties in prices and yields. Farmers' decisions follow from maximization of expected utility. Under a set of specific assumptions about farmers' utility functions, Nowshirvani shows that incorporating risk in the analysis of agricultural supply may show a negative area–price response. The natural variability of land also affects the magnitude of this response. As he says, "if the diversification of cropping is not dictated by the physical conditions of production but rather by the desire to reduce risk, stabilization schemes may sometimes be more effective policy instruments than price in bringing about area shifts among crops." He also observes that when prices and yields are negatively correlated, price stabilization leads to income destabilization, which could also lead to reducing the area devoted to the crop under consideration.

Nowshirvani does not distinguish between the prices received by farmers and prices paid for the same product. However, many of his conclusions would be strengthened by making this differentiation.

Two issues often raised are:

- Which is the relevant variable for characterizing farm supply response – acreage or farm output?
- Which price should be used – average, pre-sowing, post-harvest, modal, or another?

Several authors, including Nerlove, R. Krishna, and Narain, used acreage. Different prices have, however, been used in various studies. For example, Nerlove used an average price, while R. Krishna used post-harvest prices. Rao and J. Krishna, who examined this issue in two studies (1965, 1967), attempted to determine the impact of different prices on acreage estimations; they used a total of 21 different combinations or sets of prices in their work. It is thus difficult to conclude that any particular set of prices best explains supply response.

Whatever prices one might use, A. Parikh (1972) questions the validity of the common assumption that farmers react primarily to prices. In a static framework, he argues, prices can be the major determinant of land allocation. In a dynamic setup, however, there are often other factors, such as technological changes, that might equally influence allocation decisions. In time-series analyses, this becomes even more important. Further, when one is dealing with individual crops rather than with aggregate agricultural production, relative profitability determines the extent to which one crop is substituted for another.

A. Parikh uses relative price as well as yield expectations (though not a combined relative revenue expectation) and, in an essentially Nerlovian model, estimates Indian farmers' market responsiveness for commercial crops from data covering the period from 1900 to 1939.

4 ESTIMATIONS

Two points should be noted with respect to estimation. First, while a large number of the studies discussed in Section 3 are based on time-series data, several do not specify whether they allowed for autocorrelation. The exact form of autocorrelation in the ultimate reduced form depends on the assumptions made about the nature of the disturbance terms involved in the original model; sometimes, applying the Cochrane–Orcutt technique may not be sufficient.

Second, some studies accepted the naive expectation model as far as the price expectation functions are concerned, i.e., $P_t^* = P_{t-1}$. This is probably because of the problem of parameter identification. In some studies, P_t^* is written as a distributed lag of past prices, assuming that the lag is known.

We believe that prices cannot adequately explain acreage response and that, for most crops, revenue relative to that of competing crops is a more appropriate variable. After summarizing our experience with the traditional Nerlovian model, we separately estimate the revenue expectation functions for each crop. As we have time-series data, we employ the Box–Jenkins method to estimate these revenue expectation functions. We then use these crop revenue expectation functions in estimating the Nerlovian equations required.

4.1 Indian Crops

Rice, the most widely grown crop in India, accounted for roughly 23 percent of the total gross cropped area in the country in 1974. Wheat has gradually evolved to be the second most important crop, closely followed by jowar and then by bajra. Wheat's total gross cropped area is around 50 percent of that of rice. Other important crops are maize, gram, barley, and ragi among the food grains, and groundnut, rapeseed and mustard, sesamum, and cotton among the nonfood crops. Sugarcane accounted for 1.6 percent of the total area in 1974.

Appendix A provides data on the substitutable crops for most Indian states. Appendix B provides data on the sowing, harvesting, and peak marketing seasons of principal crops in India. (See Government of India 1967.) The inter-crop substitution pattern generally varies from state to state owing to differences in the soils and, at least to some extent, in the customs and habits of the inhabitants in different states. These factors are implicit in the sowing and harvesting periods for different crops, shown in Appendix B. To arrive at a substitution pattern for crops at the national level, the following considerations were taken into account:

- Principal and competing crops in each state
- Relative importance of each crop at the national level
- Relative importance of each state with regard to the crop at the national level
- Sowing and harvesting periods for different crops

Based on these considerations, we formulated the following overall substitution pattern of crops at the national level:

- Rice, ragi, jute, mesta, and sugarcane
- Wheat, gram, barley, and sugarcane
- Jowar, bajra, maize, cotton, oilseeds, and sugarcane
- Groundnut, rapeseed and mustard, sesamum, and other oilseeds
- Fruits, vegetables, condiments, and spices
- Rubber
- Coffee
- Tea
- Tobacco

We then classified the crops into the groups shown in Table 1.

Five points should be noted. First, crops in different groups are usually grown in different soils, seasons, or both. Sugarcane is an exception: it grows in more than one season, and when it is ratooned – that is, when the sugarcane is not planted but is allowed to grow from the stem left in the ground after the first harvest – the crop can cover more than one year.

Second, Appendix A shows that sugarcane (group 9 of Table 1) competes with most of the crops in groups 1, 2, and 3 of Table 1. However, sugarcane may not be the principal competing crop for some of these crops, and we have computed relative revenue for each crop only with respect to its two most important competing crops. Nevertheless, we did investigate the effect of increasing the irrigation facilities for sugarcane (which might increase the yield, and hence the revenue) on the acreage response of each crop in groups 1, 2, and 3.

Third, the oilseeds (group 4) compete with the crops in group 3, but group 4's total area is much smaller than that of group 3. The competition in the reverse direction may thus not be great.

Fourth, except for those mentioned in the two preceding paragraphs, no inter-group substitution possibilities are assumed to be possible at the national level.

Fifth, the residual components in the first four groups contain small millets and pulses. These do not compete to a great extent with the other crops in the respective groups.

4.2 Our Experience with the Nerlovian Model

We began our estimation exercises by applying the Nerlovian model as such. The set of variables in our analysis is as follows:

A_{igt}, P_{igt}, Y_{igt}, R_{igt} are the area, wholesale price index, yield per hectare, and rainfall index, respectively, of the ith crop in group g in period t

t refers to the time period

$*$ refers to the desired or expected values

$\Pi_{igt} = P_{igt}\, Y_{igt}$ is the revenue of the ith crop in group g

Π_{k1gt} and Π_{k2gt} are the revenues of competing crops $k1$ and $k2$

TABLE 1 Crops and groups in the system.

	Group (g)									
Crop (i)	1	2	3	4	5	6	7	8	9	10
1	Rice	Wheat	Maize	Groundnut	Fruits and vegetables	Rubber	Coffee	Tea	Sugarcane	Tobacco
2	Ragi	Gram	Bajra	Sesamum	Condiments and spices					
3	Jute	Barley	Jowar	Rapeseed and mustard						
4	Mesta		Cotton							
Q_g	Residual	Residual	Residual	Residual						
Group total[a]	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}

[a]Sum of area in all groups = total gross cropped area = $A_G = \sum_i A_{ig} + Q_g$.

I_{gt} is the total irrigated area of all crops in group g

I_{Gt} is the total irrigated area in the country

I_{st} is the irrigated area of sugarcane

For the first attempt we used the following equations for the model:

$$A^*_{igt} = a_0 + a_1 \Pi^*_{igt} + a_2 R_{igt} + a_3 \Pi^*_{k1gt} + a_4 \Pi^*_{k2gt} + U_t \quad (11)$$

$$\Pi^*_{igt} - \Pi^*_{ig(t-1)} = \beta(\Pi_{igt-1} - \Pi^*_{igt-1}) \quad (12)$$

$$\Pi^*_{k1gt} - \Pi^*_{k1g(t-1)} = \beta(\Pi_{k1gt-1} - \Pi^*_{k1gt-1}) \quad (13)$$

$$\Pi^*_{k2gt} - \Pi^*_{k2gt-1} = \beta(\Pi_{k2gt-1} - \Pi^*_{k2gt-1}) \quad (14)$$

$$A_{igt} - A_{igt-1} = \gamma(A^*_{igt} - A_{igt-1}) - U_t \quad (15)$$

where $U_t = \rho U_{t-1} + \epsilon_t$ and $0 < |\rho| < 1$.

These give a reduced form

$$\begin{aligned} A_{igt} - (1-\beta)A_{igt-1} &= a_0\beta\gamma + a_1\beta\gamma\Pi_{igt-1} + (1-\gamma)[A_{igt-1} - A_{igt-2}(1-\beta)] \\ &+ a_2\gamma[R_{igt} - (1-\beta)R_{igt-1}] + a_3\beta\gamma(\Pi_{k1gt-1}) + a_4\beta\gamma(\Pi_{k2gt-1}) \\ &- [(U_t - \rho\gamma U_{t-1}) - (1-\beta)(U_{t-1} - \rho\gamma U_{t-2})] \quad (16) \end{aligned}$$

We first assumed the price expectation coefficient to be the same for principal and competing crops. We also specified the disturbance term, which serves primarily to facilitate application of readily available techniques to account for autocorrelation. The assumption of the same price expectation coefficient for all competing crops implies that the equations for these crops should be estimated simultaneously, which was our original intention. We did make a separate estimate for each crop to observe the model's behavior, but we encountered difficulties. We estimated eq. (16), the reduced form of eqs. (11) through (15), for a range of specified values of β. We scanned the range $0 < \beta \leqslant 1$ and observed the highest $\bar{R}^2$.

We were somewhat disappointed by the results. We observed that the highest $\bar{R}^2$ was associated with $\beta = 1$ for almost all crops. The values of $\bar{R}^2$ were of course highly attractive in most cases. One could perhaps have accepted such estimates, if β were to be equal to 1.0, in some of the crops, but not in all; our estimates would then become questionable in spite of the high $\bar{R}^2$. This result does not seem to be a quirk of the estimating procedure (such as may result from the likelihood function being monotonic with respect to β) because the estimates obtained in a similar way by Cummings (1975) do not show the same rigid pattern of β always taking a corner value of the possible range.†

†When, to further explore this problem, we extended the range of β to 2.0, we obtained interior estimates of β for a number of crops.

Accepting these estimates would have meant that farmers in India have only naive expectations. However, we did not believe that this could be the case with all farmers. We could not overcome this difficulty, however, even by alternative specifications involving prices, trend variables, and logarithmic values of the variables.

Referring again to the Nerlovian price expectation formulation, we have

$$P_t^* = \beta P_{t-1} + (1-\beta)P_{t-1}^* \qquad 0 < \beta \leqslant 1 \tag{17}$$

This is a first-order difference equation. The solution of this equation is

$$P_t^* = H(1-\beta)^t + \sum_{\lambda=0}^{t} \beta(1-\beta)^{t-\lambda}P_{\lambda-1} \tag{18}$$

where H is a constant. Under certain assumptions made on initial conditions and other factors, this can be rewritten as

$$P_t^* = \sum_{\lambda=0}^{t} \beta(1-\beta)^{t-\lambda}P_{\lambda-1} \tag{19}$$

That is, the expected "normal" price is a weighted average of past prices. Supposing that the relation between actual and expected prices at period t is $P_t = P_t^* + W_t$, where W_t comprises all random shocks and disturbances,

$$P_t^* = \sum_{\lambda=0}^{t} \beta(1-\beta)^{t-\lambda}(P_{\lambda-1}^* + W_{\lambda-1}) \tag{20}$$

implies that the weights attached to the expected price value and the random disturbances are the same in each period. This obviously cannot be the case for a meaningful notion of an expectation function.

We clearly needed to formulate the revenue expectation equation differently. The presence of a secular trend in the revenues could lead to a result where β would exceed 1. If expectations reflect secular trends in relative revenues, it seems reasonable to assume that farmers observe the levels of prices and revenues over time and are also aware of any random shocks (which may be of a short-term nature) to which the variables have been subjected. The future expected price or revenue should adequately account for this process of movement and occasional random shocks.

An ARIMA model seemed to be more satisfactory:

$$\begin{aligned} P_t^* = P_t - W_t = \phi_1 P_{t-1} + \phi_2 P_{t-2} + \phi_3 P_{t-3} + \cdots + \mu + \theta_1 W_{t-1} \\ + \theta_2 W_{t-2} + \theta_3 W_{t-3} + \cdots + \end{aligned} \tag{21}$$

where P_t^* is the expected price, P_t is the actual price, W_t is the difference between them, and μ is a constant. If we compare eqs. (17) and (21) by expanding eq. (18) as

$$P_t^* = H(1-\beta)^t + \beta P_{t-1} + \beta(1-\beta)P_{t-2} + \beta(1-\beta)^2 P_{t-3} + \cdots + \tag{22}$$

we see that the Nerlovian formulation of the expectation equation is simply a special case of eq. (21) where the values of θ_1, θ_2, and so forth are all set to zero ($\theta_1 = \theta_2 = \cdots = 0$) and the other parameters are restricted to follow a geometric series. While eq. (21) implies that farmers, in formulating expectations for the future, take into account not only past realized prices but also the extent to which their expectations are off the mark, eq. (17) implies that they ignore past differences between their expectations and realizations.

4.3 Estimating Crop Revenue Expectation Functions

In this section we present the estimates of revenue expectation functions based on the Box–Jenkins methodology (see Box and Jenkins 1970). A time series constituting a discrete linear stochastic process of $\{X_t\}$ can be written as

$$X_t = \mu + \psi_0\epsilon_t + \psi_1\epsilon_{t-1} + \psi_2\epsilon_{t-2} + \cdots + \tag{23}$$

where ψ_js are the weights attached to random disturbances of different time periods. μ is a constant that determines the level of the time-series process. If a given time series is stationary, it fluctuates randomly about a constant mean; this means that the stochastic process remains invariant over time. If the time series is not stationary, it does not have a natural mean. If eq. (23) is a convergent sequence, the process is said to be stationary; if it is divergent, it is said to be nonstationary. Some nonstationary time series can be reduced to stationary series (which are then called "homogenously nonstationary," before reduction) by applying an appropriate degree of differencing d to the original series.

∇, the differencing operator, and B, the backward shift operator, are defined as follows:

$$\nabla^d X_t = (1 - B)^d X_t$$

where

$$B^n X_t = X_{t-n}$$

Then a stationary series $\{Y_t\} = \{\nabla^d X_t\}$ can be obtained from a nonstationary series $\{X_t\}$. A "parsimonious" approach toward estimation requires rewriting the sequence (eq. (23)) as an equation containing on the right-hand side only a finite number of lagged dependent variables p and moving average variables q. Box and Jenkins developed a satisfactory econometric methodology to estimate a model to forecast the value of a variable by being able to identify the stationary and random components of each of its past values. Generally, a Box–Jenkins autoregressive integrated moving average (ARIMA) process can be written for a time series $\{\Pi_t\}$ as

$$\Pi_t = \phi_1\Pi_{t-2} + \phi_2\Pi_{t-2} + \phi_3\Pi_{t-3} + \cdots + \mu + \theta_1 w_{t-1} + \theta_2 w_{t-2}$$
$$+ \theta_3 w_{t-3} + \cdots + \tag{24}$$

where w_t is the white noise or random disturbance in period t. Equation (24) is the ultimate equation to be estimated, in which the number of parameters depends on the values of p, q, and the degree of differencing d. Henceforth in this report, we indicate the ARIMA schemes that we estimate by p, q, and d, in that order. For each crop we applied the following ARIMA schemes (using an International Mathematical and Statistical Library (IMSL) computer programming package) to estimate $\Pi_{igt}(=P_{igt}Y_{igt})$ as a function of past revenues and white-noise (random disturbance) values in the form of eq. (24):

(p,q,d): (1,1,0), (1,2,0), (2,1,0), (1,1,1), (1,2,1), (2,1,1)

We selected the best of these six schemes by first, checking the stationary conditions of the estimated series, implying certain restrictions that the estimated parameter values must satisfy (parameter values can be expressed in terms of the autocorrelation function) and second, making a χ^2 test on the residual autocorrelations.

Table 2 shows the selected schemes, the results of the estimates, and the χ^2 values based on the residual autocorrelations. The numbers representing the ARIMA scheme are written in the order p, q, d, where p is the number of autoregressives, q is the number of moving averages, and d is the degree of differencing applied to make the original "homogenously nonstationary" series stationary.

Each of these estimated equations shows a stationary process of a variable for sequential values over time. The estimations provide the appropriate weights to be given for past values of the stationary and random components of a variable. Dropping the subscripts for crops, we write the farmers' expected normal revenue as

$$\Pi_t^* = \Pi_t - w_t = \phi_1\Pi_{t-1} + \phi_2\Pi_{t-2} + \phi_3\Pi_{t-3} + \cdots + \mu + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \theta_3 w_{t-3} + \cdots + \tag{25}$$

In the next section the estimated values of Π_t^* from eq. (25), subsequently referred to as $\hat{\Pi}_t$, are used in reestimating the Nerlovian model.

4.4 Estimating the Acreage Response Model

While reestimating the model, we made additional modifications to the equations presented in Section 4.2.

First, instead of treating the revenues of the principal and competing crops as separate variables, we introduced only one variable Z_{igt}, defined as follows:

$$Z_{igt} = \hat{\Pi}_{igt}/(\hat{\Pi}_{k1gt}\hat{\Pi}_{k2gt})^{1/2} \tag{26}$$

or

$$Z_{igt} = \hat{\Pi}_{igt}/\frac{1}{2}(\hat{\Pi}_{k1gt} + \hat{\Pi}_{k2gt})$$

where

$$\Pi_{igt} = P_{igt}Y_{igt}$$

TABLE 2 Box–Jenkins ARIMA process schemes and results of expectation function estimations.

Variable (π_t)	ARIMA scheme	ϕ_1	ϕ_2	ϕ_3	μ	θ_1	θ_2	ω_{1972}	ω_{1973}	ω_{1974}	χ^{2a}
Bajra price	110	0.9364			8.0810	0.7367		31.65	49.58	13.00	6.99
Bajra yield	120	0.8473			0.0547	−0.1092	−0.5128	0.452	0.332	0.540	8.21
Barley revenue	121	1.2735	−0.2735			−0.9288	1.4495	16.604	74.763	0.00	4.31
Sugarcane revenue	111	0.4641	0.5359			0.8154		284.605	−14.462	0.00	7.45
Cotton revenue	121	0.5718	0.4282			−0.4374	0.7503	−4.444	18.277	0.00	6.29
Groundnut revenue	211	0.0613	−0.0497	0.9884		0.2528		−14.014	153.892	0.00	3.77
Gram revenue	121	0.7787	0.2213			−0.2960	0.6019	71.154	−5.263	0.00	6.39
Jute revenue	121	0.6927	0.3074			0.1676	−0.3014	13.143	−68.96	0.00	5.98
Jowar revenue	121	1.6994	−0.6994			−0.3521	0.7676	36.258	44.130	0.00	5.76
Mesta revenue	120	0.8447			65.7772	−0.2742	−0.0349	89.186	10.514	105.782	9.30
Maize revenue	111	0.6019	0.3981			0.2145		61.995	123.436	0.00	5.77
Maize price	121	1.7914	−0.7914			−0.3660	0.6225	49.023	65.896	0.00	5.71
Maize yield	120	0.9719			0.0264	−0.9729	1.2282	−0.018	−0.048	−0.041	8.97
Rice revenue	111	0.8705	0.1296			0.9236		65.422	9.374	0.00	7.87
Ragi revenue	111	0.4856	0.5144			1.4122		55.297	36.927	0.00	5.02
Rapeseed and mustard revenue	211	0.0069	0.2066	0.7866		0.4297		27.818	39.435	0.00	9.12
Sesamum revenue	211	0.5887	−0.4238	0.8351		0.4254		6.685	17.001	0.00	6.67
Tobacco revenue	121	0.2405	0.7595			1.2292	0.9618	8.465	32.876	0.00	5.72
Wheat revenue	211	0.2497	0.4024	0.3480		0.7508		8.749	234.803	0.00	3.06

NOTES: $\pi_t = \phi_1\pi_{t-1} + \phi_2\pi_{t-2} + \phi_3\pi_{t-3} + \mu + \theta_1\omega_{t-1} + \theta_2\omega_{t-2} + \omega_t$.
μ = a constant equal to the mean of the series if $d = 0$.
ω_t = white noise in time t.
Degrees of freedom = number of observations (21) – number of parameters.

[a] Based on the residual autocorrelations.

Z_{igt} gives the revenue of crop i relative to competing crops $k1$ and $k2$ computed on the basis of either geometric or arithmetic average, and ($\hat{}$) denotes the estimated value obtained from the Box–Jenkins exercise.

Second, we introduced three irrigation variables: I_{Gt}, to catch the impact of further irrigation in the country; I_{gt}/I_{Gt}, to capture the effect of the share of the gth group of crops in the total irrigated area; and I_{st}/I_{Gt}, to account for the irrigated area devoted to sugarcane and thus not available for the crop being considered.

Third, we constructed the rainfall index for the crop by taking a weighted average of monthly rainfall in different states for the months critical to a crop. We used the production levels of the crops in various states as weights (see Ray 1977).

Fourth, we specified the model in a multiplicative way as follows:

$$A^*_{igt} = a_0(Z^*_{igt})^{a_1}(R_{igt})^{a_2}(I_{gt}/I_{Gt})^{a_3}(I_{st}/I_{Gt})^{a_4}(I_{Gt})^{a_5}V_t \tag{27}$$

$$Z^*_{igt} = Z_{igt} \tag{28}$$

which is defined in eq. (26) as

$$A_{igt} = (A^*_{igt})^{\gamma}(A_{igt-1})^{1-\gamma} \tag{29}$$

Substitution after taking logarithms yields the following reduced form equation:

$$\log A_{igt} = a_0\gamma + (1-\gamma)\log A_{igt-1} + a_1\gamma \log Z_{igt} + a_2\gamma \log R_{igt}$$
$$+ a_3\gamma \log (I_{gt}/I_{Gt}) + a_4\gamma \log (I_{st}/I_{Gt}) + a_5\gamma \log (I_{Gt}) + \gamma \log V_t \tag{30}$$

where $U_t = \log V_t$ is normally distributed as $N(0,\sigma^2)$.

In estimating eq. (30), several essential points should be kept in mind (see Johnston 1972).

First, as the data used represent a time series, autocorrelation is possible. In such a case, applying the ordinary least-squares (OLS) estimator would give unbiased estimates, but the sampling variances might be underestimated.

Second, the presence of the lagged dependent variable on the right-hand side (in the absence of autocorrelation) leads to estimates that are consistent but that can be biased in small samples. However, if OLS is applied in the presence of autocorrelation, the combination does not even yield consistent estimates.

Third, if the disturbance term and the dependent variable in eq. (30) are correlated, the disturbance term is also correlated with at least one explanatory variable, especially under autocorrelation (which, again, gives biased estimates in small samples).

Fourth, under such circumstances we cannot rely on the conventional Durbin–Watson test for autocorrelation. Though the presence on the right-hand side of three or four exogenous variables (such as rainfall, relative revenue, or irrigation) other than the lagged dependent variable helps to reduce the asymptotic biases of the estimates in such cases (see Malinvaud 1970), we decided to allow for autocorrelation, and we assumed a first-order autocorrelation scheme. We initially used the Cochrane–Orcutt technique in estimation.

However, we suspected that, at least in some cases, this technique might yield only a local optimum; this had been our experience in several other exercises. Hence we preferred a scanning technique to the Cochrane–Orcutt technique for estimating the autocorrelation parameter ρ in $U_t = \rho U_{t-1} + \epsilon_t$. We estimated eq. (30) for 40 values of ρ for each crop, over a range of $-1.00 \leqslant \rho \leqslant 1.0$ with a step size of 0.05, and observed the highest $\overline{R}^2$. Interestingly, however, for many crops the estimate of ρ turned out to be zero, implying that U_t and U_{t-1} are not correlated. In this case the previously mentioned problem of correlation between the disturbance term and an explanatory variable might not exist because the estimated revenue term, rather than the actual revenue term, might be one of the explanatory variables on the right-hand side.

We took most of our data from *Estimates of Area and Production of Principal Crops in India* (Government of India 1970–1976). These volumes, published yearly, cover data on area, production, yield, and irrigation area. We collected price data from the Office of the Economic Adviser, Ministry of Industrial Development and obtained rainfall data corresponding to each crop from Ray (1977). All these data cover the period from 1953 to 1974; there are thus 21 observations on each variable.

We estimated eq. (30) for some selected crops in the groups, using Norman (1977) for estimation purposes. We obtained acceptable results for rice, wheat, groundnut, sugarcane, and tobacco initially. We adopted three criteria for acceptability of results:

1. Proper signs of the various estimates
2. Levels of significance for the computed "t coefficients"
3. A high $\overline{R}^2$

For ragi, jute, mesta, gram, barley, and sesamum, the results were considered acceptable only for the areas of these crops relative to the areas of some other crops in the group. Thus we estimated the areas under ragi/rice, jute/ragi, mesta/ragi, gram/wheat, barley/wheat, sesamum/groundnut, and rapeseed and mustard/sesamum instead of the areas under ragi, jute, mesta, gram, barley, sesamum, and rapeseed and mustard. In these cases, A_{igt} in eq. (30) represents such relative areas (i.e., A_{igt} is replaced by A_{igt}/A_{jgt}, meaning the area of the ith crop relative to that of the jth crop in group g).

Tables 3a–c show the results of area estimation. For all the above-mentioned crops (i.e., jowar, bajra, maize, and cotton excepted), the coefficients of the revenue terms are positive. These are significant at the 5 percent level for jute, mesta, wheat, barley, rapeseed and mustard, sugarcane, and tobacco. This significance varies between 10 and 20 percent for rice, ragi, cotton, and sesamum. However, these coefficients for gram and groundnut were not significant, even at the 20 percent level. That groundnut acreage response to revenue was insignificant is somewhat perplexing, especially because it is a commercial crop.

The coefficients of the A_{igt-1} term, i.e., $1 - \gamma$ where γ is the adjustment parameter, can be explained as follows:

1. If $1 - \gamma$ is significantly different from zero, then γ is significantly different from one
2. If $1 - \gamma$ is not significantly different from zero, then γ is not significantly different from one

TABLE 3a Results of area estimation.

Serial number	Crop	Group	A_{igt-1}	Rainfall	Revnrtag	Revnrate	IASO / IARGROSS	IACN / IARGROSS	IARGROSS	Constant	Degrees of freedom	$\bar{R}^2$ / DW	RHO	Competing crops
1	Rice	1	0.9854 (88.89)	0.0708 (3.27)	0.0305 (1.64)		0.0592 (1.76)			-0.000002 (-1.68)	15	95.81 (1.82)	(-0.4)	Ragi Sugarcane
2	Rice	1	0.9823 (77.87)	0.0652 (2.67)	0.0314 (1.57)		0.0685 (1.79)	-0.0236 (-0.75)		-0.000003 (-1.73)	14	95.14 (1.73)	(-0.4)	Ragi Sugarcane
3	Ragi/rice	1	0.1078 (0.42)	0.1348 (0.79)	0.1167 (1.32)		0.2421 (1.05)		-0.2464 (-2.11)	-0.000003 (-0.19)	14	49.00 (2.09)	(-0.15)	Rice Jute
4	Jute/ragi	1	0.3634 (2.08)	0.1717 (1.44)		0.5694 (3.32)	1.5201 (1.60)	-0.3301 (-1.20)	0.9331 (2.18)	-11.8880 (-2.86)	13	57.21 (2.27)	(0.00)	Rice Ragi
5	Jute/ragi	1	0.3636 (2.04)	0.1880 (1.55)	0.5573 (3.20)		1.5485 (1.60)	-0.3287 (-1.17)	0.9489 (2.18)	-11.68 (-2.75)	13	55.78 (2.34)	(0.00)	Rice Ragi
6	Mesta/ragi	1	0.3091 (1.42)	0.0397 (0.27)		0.0874 (2.44)	2.180 (1.76)		1.6926 (2.72)	-17.7745 (-3.06)	14	59.16 (2.33)	(0.00)	Rice Ragi
7	Wheat	2	0.2599 (2.20)	0.0984 (5.77)	0.0678 (2.71)				0.8081 (6.70)	-1.8796 (-2.89)	15	96.36 (2.22)	(0.00)	Gram Barley
8	Wheat	2	0.2568 (2.19)	0.0995 (5.90)		0.078 (2.77)			0.8132 (6.78)	-1.7531 (-3.08)	15	96.43 (2.33)	(0.00)	Gram Barley
9	Gram/wheat	2	0.1627 (0.69)	-0.0911 (-0.82)		0.0678 (0.71)			-1.4081 (-3.50)	14.5845 (3.53)	15	92.97 (2.00)	(0.00)	Wheat Barley

NOTES: All variables are in their logarithmic form.
Figures in parentheses are the corresponding t values.
See Table 1 for the crops belonging to each group.
Revnrtag: revenue of the crop relative to that of competing crops where the revenue of competing crops is computed as a linear average.
Revnrate: revenue of the crop relative to that of competing crops where the revenue of competing crops is computed as a geometric average; see eq. (26).
IASO: irrigated area of the soil to which the group belongs (I_{gt})
IARGROSS: gross irrigated area of all crops in the country (I_{Gt}).
IACN: irrigated area of sugarcane (I_{st}).
DW: Durbin–Watson statistic.
RHO: autocorrelation parameter in $U_t = \rho U_{t-1} + \epsilon_t$.

TABLE 3b Results of area estimation.

Serial number	Crop	Group	A_{igt-1}	Rainfall	Revnrtag	Revnrate	$\frac{\text{IASO}}{\text{IARGROSS}}$	$\frac{\text{IATOSD}}{\text{IARGROSS}}$	IARGROSS	Constant	Degrees of freedom	$\frac{\bar{R}^2}{DW}$	RHO	Competing crops
10	Barley/wheat	2	0.4320 (3.09)	−0.0399 (−0.99)	0.0898 (2.34)				−0.91110 (−3.52)	8.7462 (3.42)	15	95.35 (2.37)	(0.00)	Wheat Gram
11	Barley/wheat	2	0.4388 (3.14)	−0.0388 (−0.96)		0.0924 (2.34)			−0.8985 (−3.46)	8.56 (3.32)	15	95.35 (2.38)	(0.00)	Wheat Gram
12	Cotton/maize	3	0.9008 (9.46)	0.0182 (1.36)	0.0654 (1.82)					−0.000051 (−1.84)	16	86.90 (1.63)	(0.10)	Jowar Maize
13	Groundnut	4	0.9480 (27.79)	0.0895 (1.24)		0.0465 (0.91)				0.000019 (2.53)	16	78.98 (1.88)	(−0.35)	Sesamum
14	Sesamum/ groundnut	4	0.5489 (6.37)	0.0858 (0.90)		0.0737 (1.54)				−0.7785 (1.76)	16	69.16 (2.15)	(0.00)	Groundnut Groundnut
15	Rapeseed and mustard/sesamum	4	0.3617 (2.68)	0.0787 (2.21)		0.1336 (2.17)	−0.917[a] (−1.30)	0.1974 (2.94)		0.000011 (−2.11)	14	75.29 (1.70)	(−0.6)	Sesamum Sesamum
16	Sugarcane	8	0.0949 (0.42)	−0.2296 (−1.47)	0.1989 (1.98)				0.7230 (3.83)	0.000014 (0.80)	15	68.58 (1.70)	(0.25)	Rice Wheat
17	Sugarcane	8	0.0852 (0.38)	−0.2473 (−1.58)		0.2020 (2.10)			0.7233 (3.90)	0.00014 (0.84)	15	69.27 (1.67)	(0.25)	Rice Wheat
18	Tobacco	9	0.1762 (1.45)	0.1559 (3.39)		0.1140 (4.58)				3.7282 (5.55)		75.60 (1.85)		No competing crops

NOTES: All variables are in their logarithmic form.
Figures in parentheses are the corresponding t values.
See Table 1 for the crops belonging to each group.
Revnrtag: revenue of the crop relative to that of competing crops where the revenue of competing crops is computed as a linear average.
Revnrate: revenue of the crop relative to that of competing crops where the revenue of competing crops is computed as a geometric average; see eq. (26).
IASO: irrigated area of the soil to which the group belongs (I_{gt}).
IARGROSS: gross irrigated area of all crops in the country (I_{Gt}).
IATOSD: irrigated area of total oilseeds.
DW: Durbin–Watson statistic.
RHO: autocorrelation parameter in $U_t = \rho U_{t-1} + \epsilon_t$.
[a]Proportion of the irrigated area of competing crops other than oilseeds.

TABLE 3c Results of area estimation.

Serial number	Crop	Group	A_{igt-1}	Rainfall	Expected price	Expected yield	IASO/IARGROSS	IARGROSS	Constant	Degrees of freedom	$\bar{R}^2$ / DW	RHO
19	Maize	3	0.0605 (0.25)	0.0158 (0.25)				0.7631 (3.94)	0.000004 (1.20)	16	92.49 (1.70)	(0.65)
20	Maize	3	0.552 (3.59)	0.1107 (1.61)	0.1454 (4.87)	0.1083 (0.78)			2.5942 (2.38)	15	95.73 (2.71)	(0.00)
21	Maize	3	0.4893 (2.90)	0.0882 (1.41)	0.1230 (4.27)			0.2104 (1.10)	1.1541 (1.36)	15	95.90 (2.26)	(0.00)
22	Maize	3	0.13 (0.51)	0.0192 (0.33)		-0.2021 (-1.40)		0.7060 (3.40)	0.000004 (1.38)	15	92.67 (1.77)	(0.80)
23	Bajra	3	0.1371 (0.60)	0.1803 (2.54)			0.7751 (3.72)		8.9726 (3.86)	16	55.08 (1.98)	(0.00)
24	Bajra	3	0.1697 (0.73)	0.0729 (1.19)				0.2818 (3.43)	4.5248 (2.44)	16	51.76 (1.68)	(0.00)
25	Bajra	3	0.1637 (0.73)	0.1545 (2.12)	0.0574 (1.23)	0.4742 (1.49)	0.4742 (1.49)		7.9027 (3.23)	15	56.60 (2.06)	(0.00)
26	Bajra	3	0.2678 (1.17)	0.1213 (1.60)		0.0789 (1.65)	0.5626 (2.39)		7.6335 (3.26)	15	59.74 (2.32)	(0.00)
27	Bajra	3	0.4043 (1.92)	0.0560 (0.87)	0.0773 (2.07)	0.0770 (1.49)			5.0405 (2.72)	15	56.50 (2.29)	(0.00)
28	Jowar/ maize	3	0.3065 (1.42)	0.0341 (0.69)				-0.7492 (-3.38)	8.4956 (3.29)	16	97.51 (0.00)	(0.00)

NOTES: All variables are in their logarithmic form.
Figures in parentheses are the corresponding t values.
See Table 1 for the crops belonging to each group.
IASO: irrigated area of the soil to which the group belongs (I_{gt}).
IARGROSS: gross irrigated area of all crops in the country (I_{Gt}).
DW: Durbin–Watson statistic.
RHO: autocorrelation parameter in $U_t = \rho U_{t-1} + \epsilon_t$.

The first factor implies that farmers could not achieve their desired acreage levels immediately but could adjust their acreage to some extent. The second implies that they could adjust their acreage to the desired levels. For rice, $1 - \gamma$ is significantly* different from zero and almost equal to one, which means that rice farmers could adjust their acreage to the desired levels slowly. As rice is already the most important crop in India, accounting for 23 percent of the total, and as difficulties are involved in bringing more area under cultivation, this is understandable.

Jute, wheat, cotton, groundnut, sesamum, and rapeseed and mustard also exhibit the same phenomenon, but the adjustment parameter γ is not as low as it is for rice. For ragi, mesta, gram, sugarcane, and tobacco, this coefficient is not significant.

Except in the case of sugarcane, gram, and barley, the coefficient of rainfall is always positive. As far as irrigation is concerned, a positive coefficient of I_{gt}/I_{Gt} indicates substitution of the particular crop for the areas of the competing crops in that group, while a negative coefficient indicates that as irrigation facilities for that group increase, other crops are preferred. This argument can be extended with respect to the coefficient of I_{Gt}, which indicates the effects of increasing the total irrigated area in the country on the area devoted to the particular crop. I_{Gt} is included as a variable because many irrigation facilities in India are storage schemes permitting the transfer of water across seasons and regions, i.e., across our groups. Moreover, irrigation schemes in India are designed for extensive rather than intensive irrigation. The fluctuations in irrigation availability due to rainfall fluctuation can be significant. The sign of the coefficient I_{st}/I_{Gt} indicates the substitution trends between the crop under consideration and sugarcane.

Maize, jowar, and bajra were not included in the preceding discussion because a separate analysis, with a different hypothesis, was required for these crops. When the model as presented above was applied to these crops, our estimation results showed consistently negative and significant coefficients for the revenue variable. The $\overline{R}^2$ values were also satisfactory for all the crops; in fact, they were quite high for maize.

We considered this result to be plausible, as these three crops are primarily subsistence crops. If these crops are grown primarily for self-consumption, then farmers need only a fixed output in a given period; they adjust area allocation only to produce that output. If the productivity of the land is increased through technological or other factors, then they need to allocate less area to produce the same output; hence an increase in the yield of these crops should have a negative effect on the acreage response. However, an increase in the price of these grains leads to a positive acreage response because the farmers would then like to grow more for sale. Under these circumstances, the net effect on the revenue per acre, which is price multiplied by yield, may be a negative acreage response.

More formally, if the calorie content, yield, harvest, and market prices are defined by C, Y, P^{h}, and P^{m}, respectively; and if subscripts c and r refer to coarse grain and to rice, then $\mathrm{d}A/\mathrm{d}y < 0$ and $\mathrm{d}A/\mathrm{d}p > 0$ is possible if three conditions are met:

$$C_c A_c Y_c > C_r A_r Y_r$$

$$P^m_c A_c Y_c > P^h_r A_r Y_r$$

$$P^h_r A_r Y_r > P^h_c A_c Y_c$$

*Hereafter, significance is judged at the 5 percent level.

Imposing the first condition ensures that the farmer gets more calories from his land from coarse grain than from rice; imposing the second, that growing rice for sale to buy coarse grain is uneconomical; and imposing the third, that it is better to grow rice than coarse grain for sale.

We tested this hypothesis by dropping the revenue variable from the model and substituting yield and price variables, both separately and together. For this purpose, we used the Box–Jenkins analysis separately for the yield and price variables of these crops to estimate expected values. Tables 2 and 3a–c show the results.

The results for maize support the plausibility of the hypothesis, and the $\overline{R}^2$ values range from 92 percent to 96 percent. Numbers 21 and 22 in Table 3c indicate that for maize $dA/dp > 0$ when $dA/dy = 0$, and $dA/dy < 0$ when $dA/dp = 0$. However, no. 20 in the same table introduces both price and yield terms; the coefficient for the yield term is not significantly different from zero, which may be due to multicollinearity between price and yield. Thus no. 20 may not be regarded as refutation of the hypothesis. While the analysis of bajra does not seem to support this hypothesis so clearly, the estimations based on price and yield variables were far better than those based on the revenue variable. Hence only these were included and are presented here.

We discovered similar findings for jowar, except that in this case, only relative area with respect to maize gave good results, and including revenue, price, or yield gave no better results than that shown in Tables 3a–c.

As previously mentioned, we did not analyze acreage response for groups 5, 6, 7, and 8, which contain fruits, vegetables, condiments, and spices; rubber; coffee; and tea, respectively. We estimated acreages of these crops merely as a percentage of the country's total gross cropped area, and we do not include estimation results for them in this report.

5 VALIDATION EXERCISES AND RESULTS

To determine the extent to which the estimated equations of crop revenue expectation and acreage response can be relied on for future projections, we decided to carry out simple validation exercises. In this section we give details of these exercises.

5.1 Crop Revenue Expectations

In this part of the exercise we simply compared the estimated values of the expected revenue, price, and yield of different crops obtained in Section 4.3 with the actual past values of these variables. These values for each crop were then plotted separately; Fig. 1 shows the plots, which correspond to the estimated equations presented in Table 2.

From these plots we can see that the estimated expected values (based on the stationary and random components of previous values) closely follow the actual values. In this respect, the performance of the estimated equations seems to be good, especially for bajra (price and yield), maize (revenue, price, and yield), rice, ragi, wheat, and tobacco. The results are also satisfactory for other crops, with the exception of groundnut, jute, and mesta, for which the expected values deviated from actual ones for many observations. This may be because in India international prices affect the prices of these crops to a greater extent than they affect the prices of other crops. It also explains the relatively unsatisfactory result obtained for acreage response for groundnut (see Section 4.4).

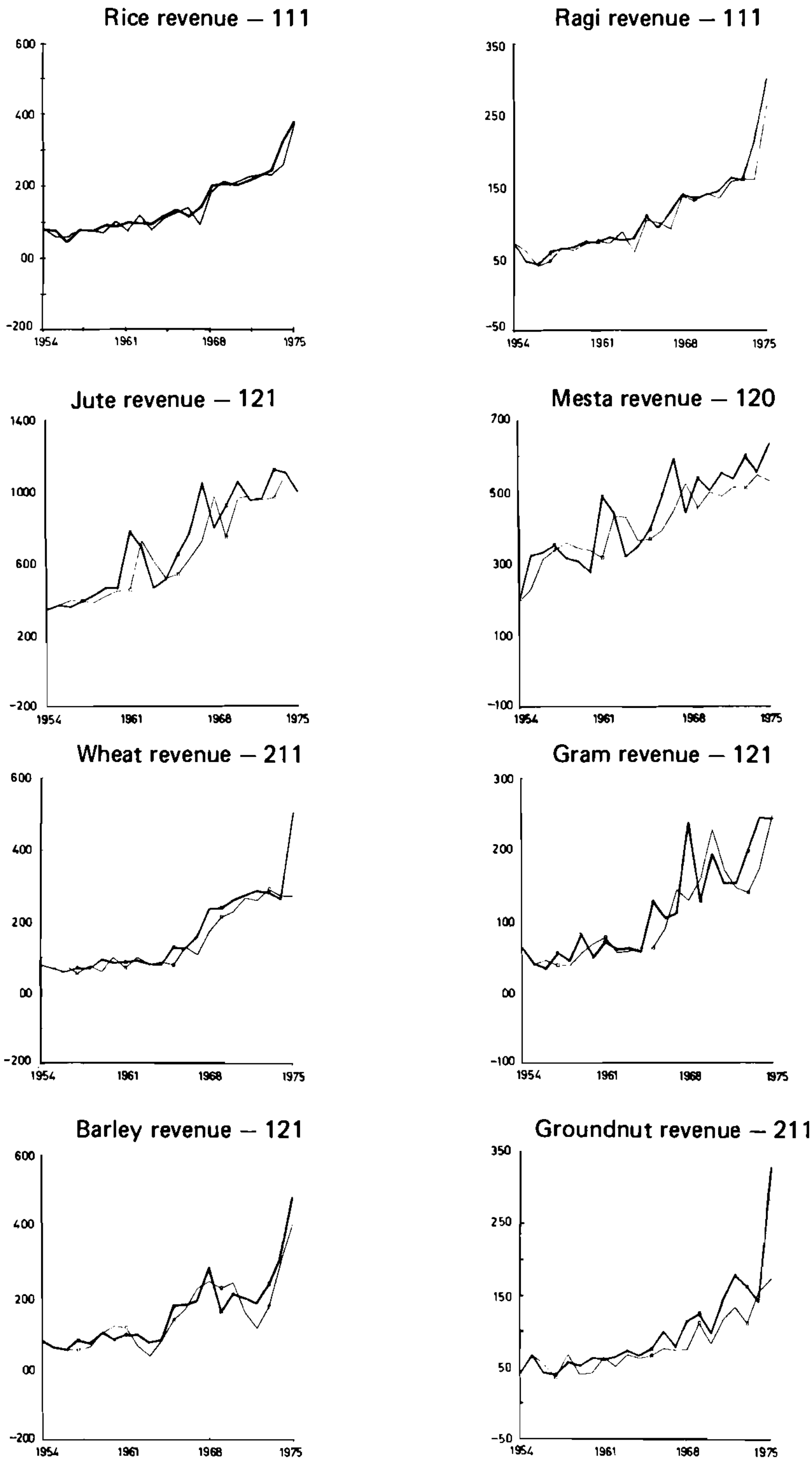


FIGURE 1 *Continued facing.*

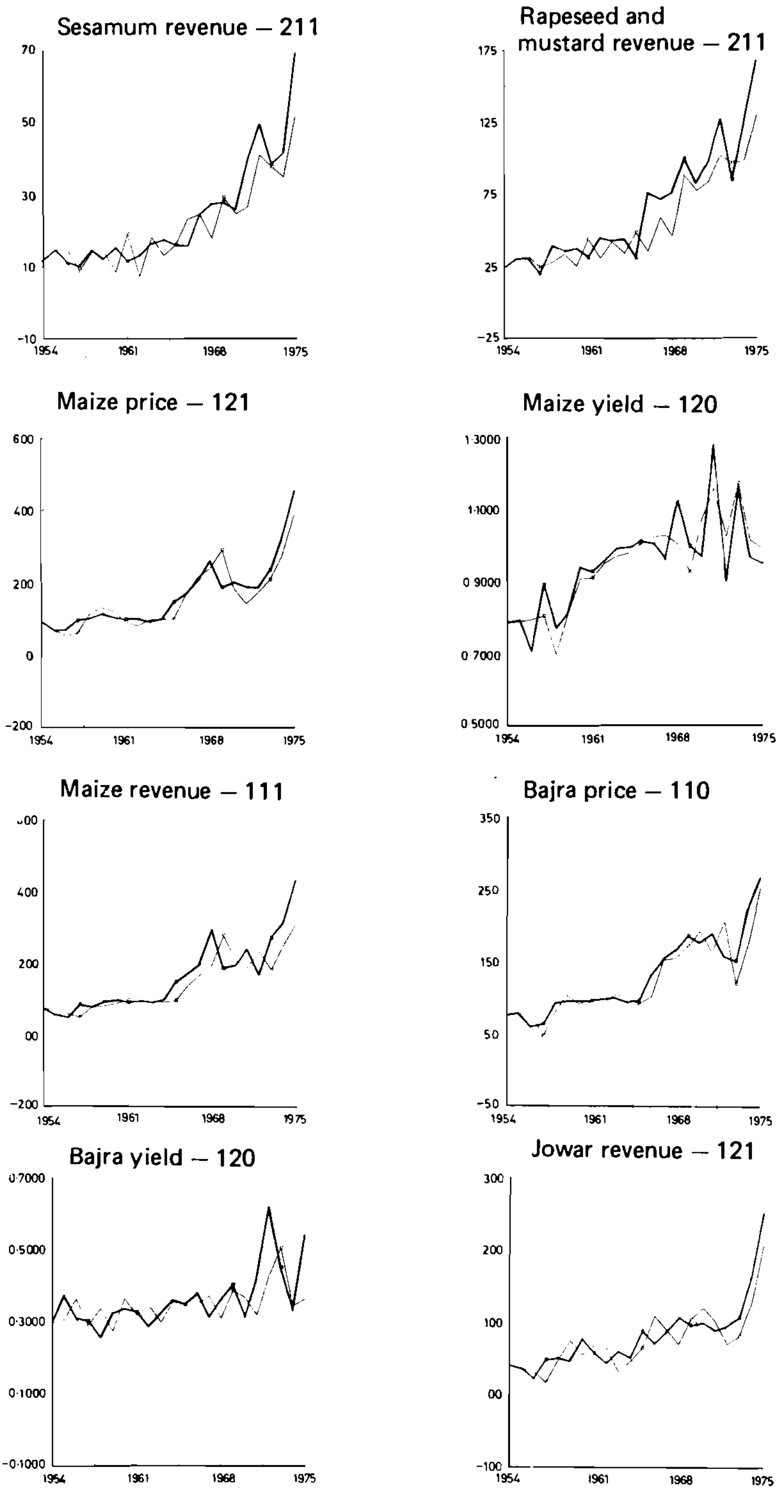


FIGURE 1 *Continued overleaf.*

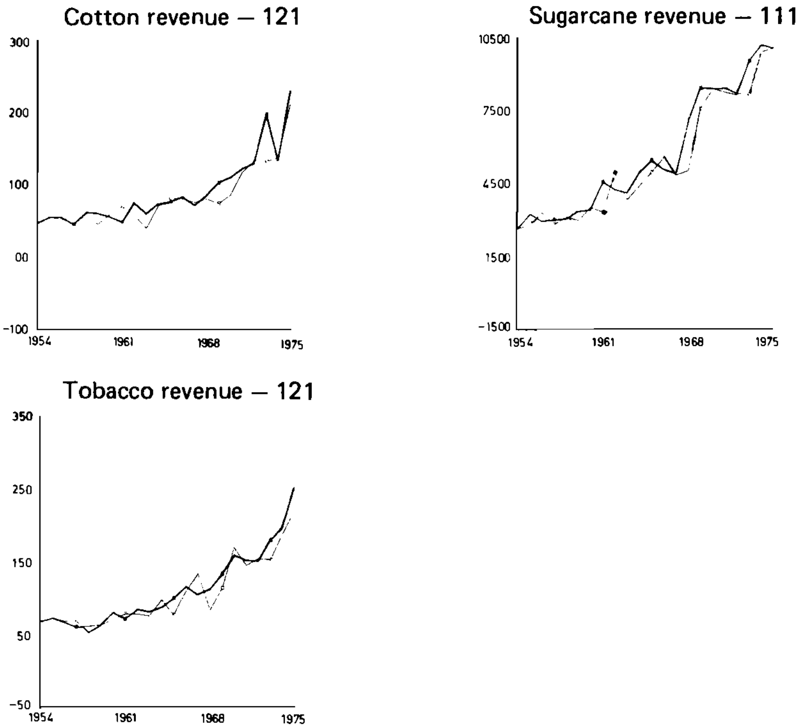


FIGURE 1 Expected (——) and actual (——) values of revenues, prices, and yields. Expected values are obtained from Box–Jenkins estimations. Numbers following crop names refer to the estimated ARIMA scheme represented by p, q, and d (see Section 4.3), where p is the number of autoregressive terms, q is the number of moving-average terms, and d is the degree of differencing. Revenues are products of wholesale price indexes and yields. Prices are wholesale price indexes, with 1961 = 100. Yields are in kg/hectare.

5.2 Acreage Response

As one of our major purposes was to use the allocation model for projection purposes in a year-by-year simulation model, we carried out a validation exercise to observe the model's behavior when it is used for a previous period. A validation exercise carried out over the period of estimation may seem to be just a look at the residuals of individual regressions. In our case, however, area projection for most crops would involve sequential use of a number of equations that were estimated separately. This projection may thus give results different from those indicated by the residuals, and a validation exercise may be required. Moreover, apart from the size of the errors, it is interesting to see to what extent the projections capture turns (ups and downs) in the data.

We estimated eq. (30) for each crop using actual data for all variables except the revenue variable, for which we obtained the numbers from the Box–Jenkins analysis. The right-hand side of eq. (30) contains as one of the variables the proportion I_{gt}/I_{Gt} of irrigated area of group g in the total irrigated area of the country and the proportion I_{st}/I_{Gt} of irrigated area of sugarcane.

Naturally, when this equation is used for future projections, one cannot have the actual values of the variables on the right-hand side, which must first be projected. Then the projected values can be inserted in eq. (30). With respect to revenue, the estimated equations of crop revenue expectation functions obtained in Section 4.3 serve the purpose. As rainfall in India has not been found to be predictable, one can only expect that it would be normal or use a sequence of rainfall, drawn as a random sample from past observations, for the future, i.e., $R_{igt} = \overline{R}_{igt}$ for the crops grown during the rainy season. For crops of the previous monsoon season, rainfall may be considered to be known.

To determine the values of the irrigation variables that appear on the right-hand side, we decided to estimate separately the proportion I_{gt}/I_{Gt} of irrigated area of every group in the country's total irrigated area.

The values obtained from these estimations were used to carry out the validation exercise. While these estimations are carried out, however, the sum total of all these proportions added over different groups in the system should be one. Hence we estimated the following sets of equations simultaneously with a constraint equation toward the additivity:

$$\sum_{g=1}^{6} (I_{gt}/I_{Gt}) + V_{st} = 1 \tag{31}$$

$$I_{gt}/I_{Gt} = a_1 + a_2 R_{gt} + a_3(I_{gt-1}/I_{Gt-1}) + a_4(I_{Gt}) + V_{gt} \qquad g = 1,6 \tag{32}$$

g = 1 for the rice group, 2 for the wheat group, 3 for the jowar group, 4 for oilseeds, 5 for sugarcane, and 6 for all other crops. R_{gt} is the rainfall index for group g (we used the rainfall index of the main crop in that group, namely, the rainfall index of rice for group 1, and so forth). Other variables are as defined in Section 4.2.

Equation (32) expresses the proportion of irrigated area of group g in the total irrigated area as a function of predetermined variables, namely, the previous year's proportion, current year's rainfall, and currently available total irrigated area. Note that I_{Gt} is generally specified from outside the system. Hence use of the scheme behind eq. (32) for projection poses no problem.

We estimated eqs. (31) and (32) simultaneously as a nonlinear least-squares problem, using the computer programming package developed by Günther Fischer at IIASA for estimation purposes; Table 4 shows the results. The estimations correspond to the minimized sum of squares of the composite residual terms $(\Sigma V_{gt} + V_{st})$. A first-order autocorrelation scheme was also imposed on each individual disturbance term V_t.

When inserted in eq. (30), the estimated values obtained for the revenue (and price and yield, as the case may be) and irrigation variables (obtained from the Box–Jenkins equations and eq. (32), respectively), yield the projected values of the acreage response. In the validation exercise we compared these projected values with the actual values. Figure 2 shows the corresponding plots, which correspond exactly to the serial numbers presented in Tables 3a–c. The ultimate results are promising, with the expectation values and actual values falling within a close range. This performance of the estimated equations seems to be especially good for rice, wheat, maize, barley, and gram. Even for the other crops, the estimated equations perform the prediction exercise satisfactorily.

However, for some crops, such as rice and sugarcane, when sudden dips or abnormal rises in actual acreage occur in one year, the expected values for the corresponding year

TABLE 4 Results of estimation of irrigation area by groups.

Irrigation area of the group containing	a_1	$a_2 \times 10^2$	a_3	$a_4 \times 10^4$	ρ
Rice and other crops	-0.0176	0.0534	0.9666	-0.0081	-0.4486
Wheat and other crops	0.0119	-0.0069	0.8949	0.0086	0.0604
Jowar and other crops	0.0541	-0.0201	0.6372	-0.0002	-0.2832
Oilseeds	-0.0092	0.0020	0.5848	0.0049	-0.2507
Sugarcane	0.0355	-0.0083	0.4820	0.0001	0.1953
All other crops	0.0791	-0.0080	0.6066	-0.0064	-0.3498

NOTES: $\sum_{g=1}^{6} (I_{gt}/I_{Gt}) + V_{st} = 1.$

$(I_{gt}/I_{Gt}) = a_1 + a_2 R_{gt} + a_3 (I_{gt-1}/I_{Gt-1}) + a_4 I_{Gt} + V_{gt} \qquad g = 1,6.$

$V_{it} = \rho V_{it-1} + \epsilon_t \qquad i = g(1 \text{ to } 6) \text{ and } s \qquad -1 \leq \rho \leq 1.0.$

The estimates correspond to the minimized sum of $(\Sigma V_{gt} + V_{st})^2$.
The following are the estimated values of $\Sigma(I_{gt}/I_{Gt})$ for different time periods: 1.0016, 0.9975, 1.0000, 1.0056, 0.9929, 1.0064, 1.0004, 0.9957, 0.9999, 1.0016, 0.9940, 1.0026, 1.0063, 1.0032, 1.0002, 0.9953, 0.9930, 1.0033, 0.9980, 1.0003, 1.0036, 0.9991, 0.9999.

(as well as the next one or two years) differ widely from the actual values because only the acreage of the previous year is present among the explanatory variables. If there is a sudden dip in the acreage in the previous year, this abnormal value of the acreage, which accounts neither for the general level nor for the possibility of recovery, is given undue weight in predicting the current year's value. If we had considered a weighted average of the acreage of a few previous years, instead of just the previous year's acreage (A_{igt-1}), by appropriately reformulating eq. (29), the acreage adjustment equation, or eq. (8), the ultimate result would have been much better.

6 POLICY IMPLICATIONS

In some planning models, demand projections are obtained by estimating an independent subsystem of demand equations, which does not form an integral part of the entire planning exercise. When the target output levels and demand projections do not match, one assumes that suitable policy measures can be devised to make them consistent. Depending on the circumstances, such measures can include adjusting relative prices of outputs, inputs, or both; adjusting taxes, subsidies, and so forth; expanding irrigation facilities; and imposing quotas on fertilizer availability. There is no guarantee, however, of the availability of a set of reasonable policies that can make the demand or supply targets achievable.

We applied the estimations reported here (see Narayana and Parikh 1979) to identify the agricultural policies implicit in the draft sixth five-year plan of the Planning Commission of India (see Government of India 1978). Based on certain assumptions about irrigation, rainfall, and so forth, we computed for rice, wheat, and their main competing crops implied relative revenues that should prevail if the targeted output levels as specified for 1982–1983 were to be realized. We then compared these implied values with the actual values during the preplan period. We found that maintaining the relative revenue of rice at approximately its present value could lead farmers to produce the targeted levels of rice output. However, we found the relative revenue of wheat that would be consistent with the targeted output

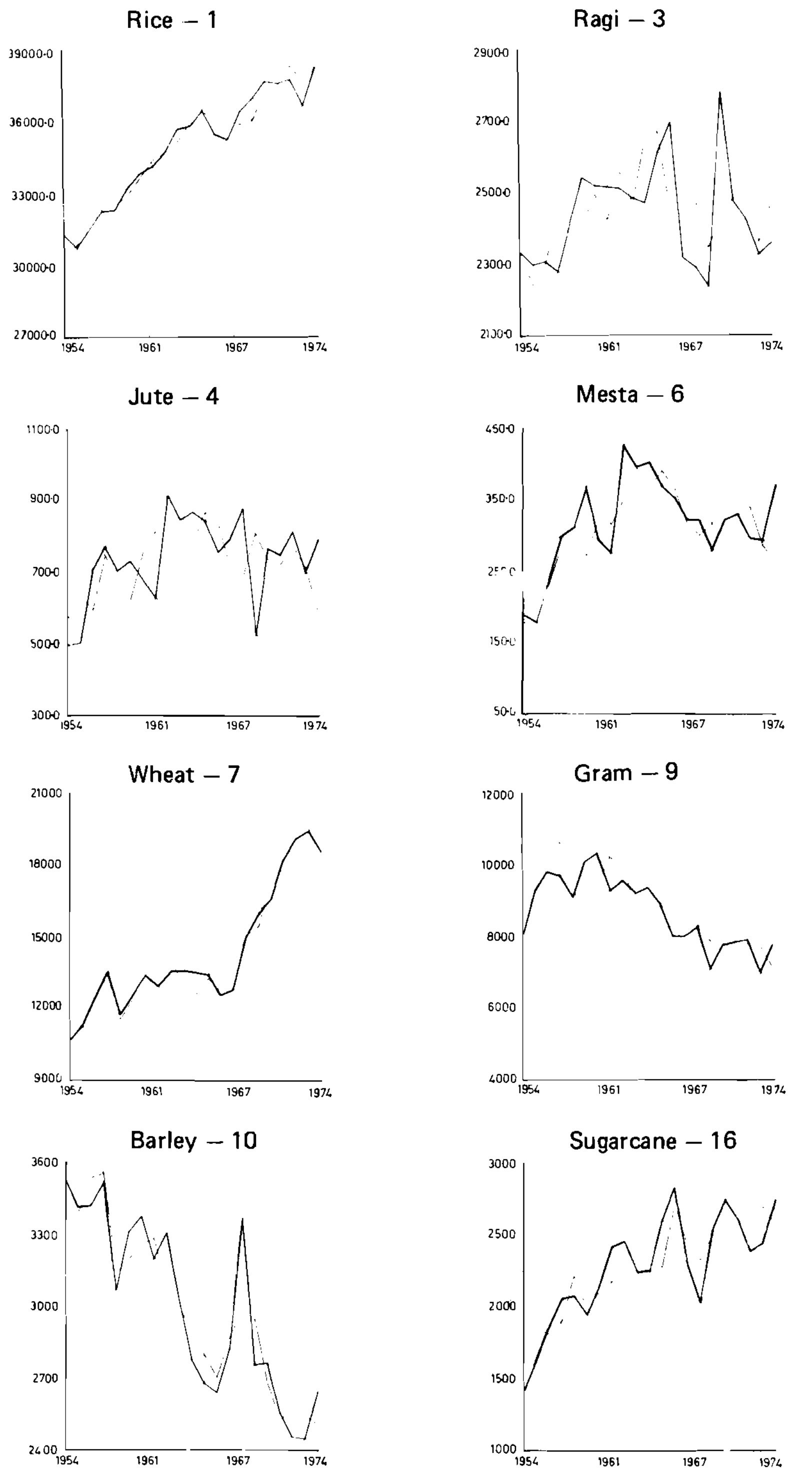


FIGURE 2 *Continued overleaf.*

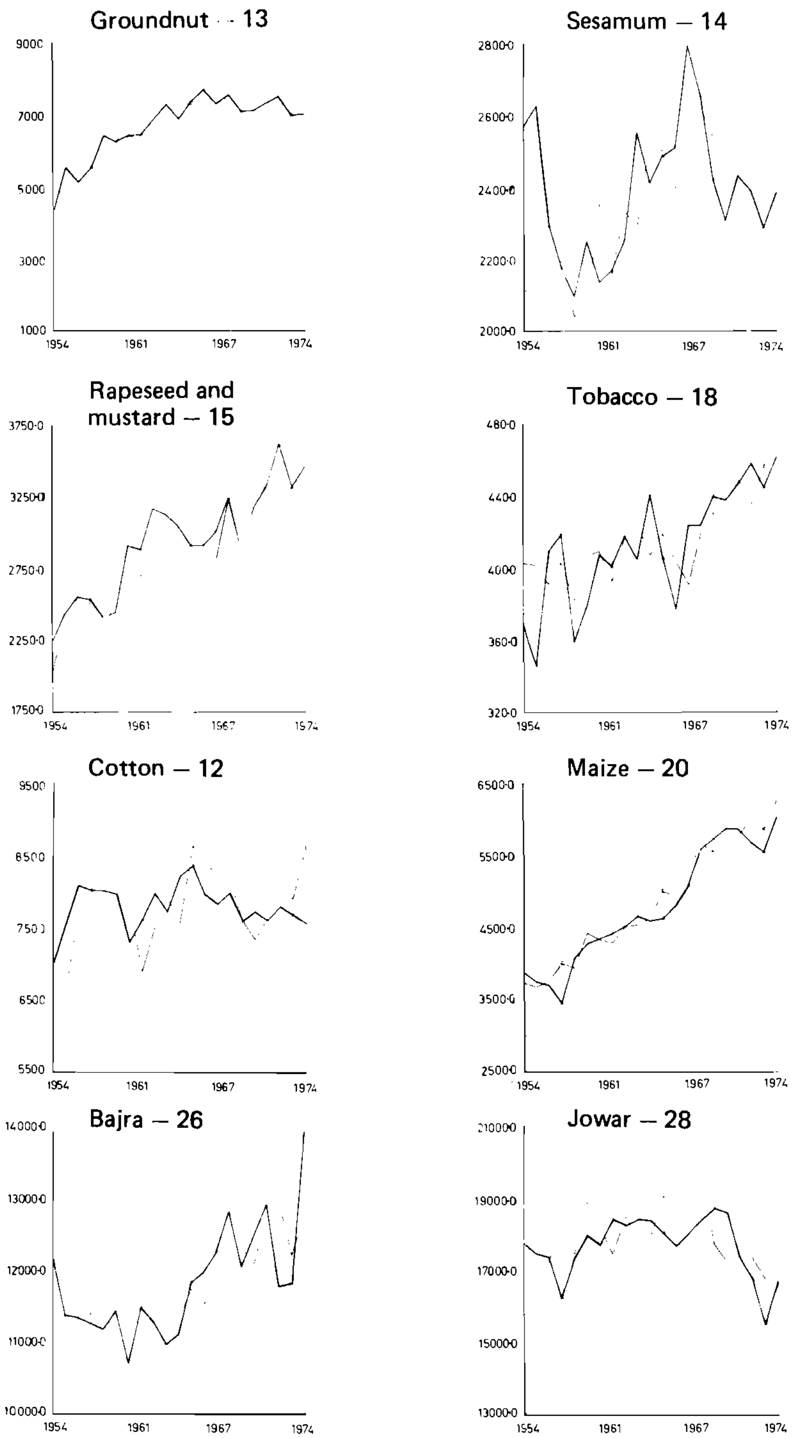


FIGURE 2 Actual (——) and projected (– –) areas (000 hectares). Projected values are obtained using projected values of predetermined variables in the right-hand side of equations in Tables 3a–c. Numbers following crop names refer to the serial numbers in Tables 3a–c.

of wheat to be an order of magnitude lower than values in the recent past. As such a change in relative revenues may be considered unlikely, this indicates that much more wheat than targeted, and much less gram and other crops that compete with wheat, is likely to be produced.

7 CONCLUSIONS

In this report we sought to model the land allocation decisions of Indian farmers. We believe that rational farmers maximize their utility within the context of opportunities, uncertainties, and risks. They cannot be expected to be insensitive to changing prices and profitabilities. We estimated acreage response for different crops using expected revenue instead of expected prices as a proxy for expected profits.

We reviewed available approaches to estimating acreage response and noted the influence of the Nerlovian model, which is based on adaptive expectations and adjustment schemes. The basic scheme behind the Nerlovian model is quite general and may be applied to the study of acreage response behavior even in developing economies, such as that of India. However, this model seems to involve a serious error of specification with respect to the formulation of the price expectation function.

A better approach to formulating an appropriate revenue (or price, as the case may be) expectation function is to identify clearly the stationary and random components involved in past values of the variable and then to attach appropriate weights to these components while predicting future values. Nerlovian specification of the expectation function cannot identify these components and thus attaches the same significance to them.

The use of Box–Jenkins methodology in estimating the crop revenue expectation functions and the subsequent use of these estimates of expected revenues in the Nerlovian adaptive acreage response model gave satisfactory results. Finally, we subjected the estimated equations to a validation exercise to judge to what extent they might be relied on for incorporation into large-scale system studies.

ACKNOWLEDGMENTS

We are deeply indebted to Michiel Keyzer for having made our "interest" rate in this exercise very high. We greatly benefited from our discussions with him, held anywhere he could be found, although we were not able to incorporate all of his suggestions here.

Klaus Frohberg and Günther Fischer helped us at several stages of this work; in return, we wish that we could blame them for at least some of the errors that may remain, but it is customary to claim that all remaining errors are ours, and we do so. We thank H. L. Chandok for providing us with important data. Sudhir D. Chitale and Frank Latko also helped us in obtaining data and in entering them on computer files. Special thanks are due to the secretaries of the Food and Agriculture Program at IIASA for patiently bearing the strain of typing the manuscript.

REFERENCES

Askari, H. and J. T. Cummings (1976) Agricultural Supply Response: A Survey of the Econometric Evidence. New York: Praeger Pub.

Behrman, J. R. (1968) Supply Response in Underdeveloped Agriculture: A Case Study of Four Major Annual Crops in Thailand 1937–1963. Amsterdam: North Holland Pub. Co.

Box, G. E. P. and G. M. Jenkins (1970) Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, Inc.

Cummings, J. T. (1975) The supply responsiveness of Indian farmers in the post-independence period. Indian Journal of Agricultural Economics 30:25.

Fisher, F. M. and P. Temin (1970) Regional specialization and the supply of wheat in the United States, 1867–1914. Review of Economics and Statistics 52(2):134–149.

Government of India (1967) Indian Crop Calendar. New Delhi: Ministry of Food and Agriculture, Community Development and Cooperation.

Government of India (1970–1976) Estimates of Area and Production of Principal Crops in India. New Delhi: Ministry of Food and Agriculture, Directorate of Economics and Statistics.

Government of India (1978) Draft Five-Year Plan, 1978–1983. New Delhi: Planning Commission.

Johnston, J. (1972) Econometric Methods. New York: McGraw-Hill Book Co.

Krishna, J. and M. S. Rao (1967) Dynamics of acreage allocation for wheat in Uttar Pradesh: A study in supply response. Indian Journal of Agricultural Economics 22, Jan.–Mar.:37.

Krishna, R. (1963) Farm supply response in India -- Pakistan: A case study of the Punjab region. Economic Journal 73, Sept.:477.

Krishnan, T. N. (1965) The marketed surplus of foodgrains: Is it inversely related to price? Economic Weekly 17, Feb. Annual number.

Lipton, M. (1966) Should reasonable farmers respond to price changes? Modern Asian Studies 1, Jan.:1.

Madhavan, M. C. (1972) Acreage response of Indian farmers: A case study of Tamil Nadu. Indian Journal of Agricultural Economics 27, Jan.–Mar.:67.

Malinvaud, E. (1970) Statistical Methods of Econometrics. Amsterdam: Elsevier–North Holland Pub. Co.

Narain, D. (1965) The Impact of Price Movements on Areas Under Selected Crops in India, 1900–1939. Cambridge: Cambridge University Press.

Narayana, N. S. S. and K. S. Parikh (1979) Agricultural planning and policy in draft sixth plan: Will farmers fulfil planners' expectations? Economic and Political Weekly XIV, Aug.:1295--1306. Special number.

Nerlove, M. (1958) The Dynamics of Supply: Estimation of Farm Supply Response to Price. Baltimore, Maryland. Johns Hopkins University Press.

Nerlove, M. (1971) Analysis of Economic Time Series by Box–Jenkins and Related Techniques. Report 7156. Chicago: University of Chicago, Center for Mathematical Studies in Business and Economics.

Norman, M. (1977) Software Package for Economic Modelling. RR-77-021. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Nowshirvani, V. F. (1971) Land allocation under uncertainty in subsistence agriculture. Oxford Economic Papers 23(3):445.

Parikh, A. (1972) Market responsiveness of peasant cultivators: Some evidence from prewar India. Journal of Development Studies 8(2):291–306.

Parikh, K. S. (1977) A Framework for an Agricultural Policy Model for India. RM-77-59. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Rao, M. S. and J. Krishna (1965) Price expectation and acreage response for wheat in Uttar Pradesh. Indian Journal of Agricultural Economics 20, Jan.–Mar.:20.

Ray, S. K. (1977) Variations in Crop Output. Delhi: University of Delhi. Institute of Economic Growth.

APPENDIX A Substitutable Crops in India

State	Crops
Andhra Pradesh	Rice, ragi, mesta Jowar, maize, bajra Cotton, groundnut, sesamum Wheat, gram
Assam	Rice, jute Moong, gram, urad, cotton, wheat
Bihar	Ragi, rice, jute Wheat, barley, peas, gram, sugarcane
Maharashtra	Linseed, wheat, gram Sugarcane, wheat, gram Jowar, bajra, maize, cotton
Madhya Pradesh	Linseed, wheat, gram Jowar, bajra, maize, cotton
Madras	Rice, ragi, mesta Jowar, maize, bajra Cotton, groundnut, sesamum
Mysore	Rice, ragi Jowar, sugarcane Cotton, groundnut Bajra, maize
Orissa	Rice, ragi, jute
Punjab	Wheat, barley, gram, peas Jowar, bajra, maize, cotton, sugarcane
Rajasthan	Jowar, bajra, maize, pulses Wheat, barley, gram, peas
Uttar Pradesh	Wheat, barley, gram, peas Jowar, bajra, maize, sugarcane
West Bengal	Autumn rice, jute Sugarcane, jute Sugarcane, rice
Delhi	Gram, wheat Wheat, barley Barley, gram
Himachal Pradesh	Wheat, barley Wheat, gram Barley, gram Wheat, mustard Maize, sesamum Maize, pulses
Manipur	Wheat, peas, mustard Maize, soyabean, sugarcane

APPENDIX B Sowing, Harvesting, and Peak Marketing Seasons of Principal Crops in India

Season	Rice (winter)	Rice (autumn)	Rice (summer)	Wheat	Jowar (kharif)	Jowar (rabi)	Bajra
Sowing	June–Oct.	Mar.–Aug.	Nov.–Feb.	Sept.–Dec.	Apr.–Aug.	Sept.–Dec.	June–Aug.
Harvesting	Nov.–Apr.	June–Dec.	Mar.–June	Feb.–May	Sept.–Jan.	Jan.–Apr.	Sept. Dec.
Peak marketing	Dec.–May	Sept.–Dec.	Apr.–July	Apr.–June	Nov.–Jan.	Feb.–Apr.	Nov.–Jan.

Season	Maize (kharif)	Maize (rabi)	Ragi	Barley	Gram	Tur (kharif)	Sugarcane
Sowing	June–Aug.	Oct.–Dec.	May–Nov.	Oct.–Dec.	Sept.–Dec.	May–Aug.	Dec.–May
Harvesting	Aug.–Nov.	Jan.–Apr.	Sept.–Mar.	Feb.–May	Feb.–May	Nov.–Apr.	Oct.–Apr.
Peak marketing	Oct.–Dec.	Mar.–Apr.	Nov.–Mar.	Apr.–June	Apr.–June	Feb.–June	Dec.–Apr.

Season	Tobacco	Groundnut	Castor	Rapeseed and mustard	Linseed	Sesamum
Sowing	July–Dec.	May–Aug.	June–Oct.	Sept.–Nov.	Sept.–Nov.	May–Sept.
Harvesting	Jan.–May	Sept.–Jan.	Oct.–Apr.	Jan.–Apr.	Jan.–May	Aug.–Dec.
Peak marketing	Feb.–June	Nov.–Jan.	Mar.–June	Mar.–May	Mar.–June	Nov.–Feb.

Season	Sesamum (rabi)	Cotton	Jute	Sannhemp	Potato (winter)	Potato (summer)
Sowing	Dec.–Feb.	Mar.–Sept.	Feb.–July	Apr.–Aug.	Aug.–Dec.	Feb.–July
Harvesting	May–Aug.	Sept.–Apr.	July–Nov.	Sept.–Jan.	Jan.–May	May–Dec.
Peak marketing	May–Aug.	Nov.–Mar.	Aug.–Jan.	Dec.–Feb.	Feb.–May	Oct.–Mar.

ECONOMIC EVOLUTIONS AND THEIR RESILIENCE: A MODEL

Manfred Breitenecker and Hans-Richard Grümm
International Institute for Applied Systems Analysis, Austria

SUMMARY

This report designs a highly aggregated macroeconomic model that can be formulated in terms of a system of ordinary differential equations (i.e., a "dynamical system"). The report consists of two parts supplementing each other in a sort of symbiosis. One part is the abstract structure of the equations — that is, the individual dependence of the time variations of the state variables (which span the state space) on the variables themselves (which in this model are E, K, and L). The other part is the parameter space, each point of which is a set of parameter values that have a well-defined economic meaning and thereby endow the system with economic content.

A particular economy is then defined by a particular point in parameter space, together with a particular point in state space (describing the "status quo") from which it evolves deterministically in time along its trajectory.

The model is analyzed carefully with the help of methods from differential topology. The following questions are answered:

- *Are there points of stationary growth in state space? If so, where are they located?*
- *What is the qualitative behavior of such a point? Is it attractive (stable) or not?*
- *Which regions of state space are slack-free — that is, describe a "desirable" economy?*
- *What is the influence of a change in the system parameters on the global behavior of a trajectory or, more generally, on the phase portrait as a whole (i.e., the set of all trajectories, roughly speaking)?*
- *Where are the regions in parameter space within which the system shows similar global behavior? In particular, where are the economic niches (regions) for which the system is globally stable?*
- *What effects do delivering and receiving investment goods (e.g., granting or receiving foreign aid) have on the qualitative behavior of the economy? To what extent can a transition out of or into a more suitable economic niche be induced by foreign aid? Similarly, what influence does the price of imported primary energy, which must be paid primarily through investment goods, have on the economy?*

As the parameter space is high-dimensional, some essential parameters have been merged into what we call scenario variables. To a certain extent, these variables reflect a particular scenario: highly or less effective use of energy, conventional or new technology in energy production, high or low emphasis on the consumption sector. The economic niches have also been determined within this scenario space.

We considered as a particular application a coupling of two economies with different qualitative behaviors (the one being within, the other outside, a stable economic niche) via foreign aid and under the influence of the price of imported energy. This led to determination of an upper limit for the price of energy.

This work is experimental; we do not intend to present a model that is in any sense final. Rather, we would like to examine more thoroughly what structural stability means, particularly with respect to long-term problems (those with a time horizon of, say, 50 years). More work and additional contributions are obviously needed.

---

1 INTRODUCTION

Both within the Energy Systems Program at the International Institute for Applied Systems Analysis (IIASA) and elsewhere, efforts have been under way to understand and therefore to conceptualize possible evolutions of energy and other systems over a long period of time -- say, 50 years. Such a time horizon is longer than that which can be treated meaningfully by the normal techno-economic models available. In the case of input–output modeling, for instance, the evolution of the input–output coefficients over time must be known if the technique is to be used purposefully. The same applies to elasticities and other input parameters in the case of econometric models. The methodology of modeling a 50-year evolution for the purpose of conceptualization is therefore new, difficult, and specific. In *Energy in a Finite World: A Global Systems Analysis* (1981), the Energy Systems Program Group at IIASA outlines in detail one approach to this problem.

Understanding such evolutions in minute detail is not always the major problem; most often, the concern is stability, and, more precisely, the stability of underlying structures. In the context of the energy problem, which is in the forefront here, a good example is the price of oil and its impact, not only on a particular economy (either importer or exporter), but also on world trade as a whole – that is, on the overall structure of economic interactions. Will there be collapses or evolutions that inherently lead to distortions? Significantly, such a question is of a holistic nature. This approach focuses on the structure of evolutions in time (and possibly in space) as a whole, not on the summations of yearly increments. The issue is thus one of structural stability.

Capital costs of new energy technologies were a special concern of IIASA's Energy Group. Since such costs tend to be high, one may wonder whether energy still works for the economy or whether the economy works for energy. While this has recently become less of a concern, it was once an important point that led to the evaluation of certain new energy technologies against the background of the rest of the economy. Furthermore, the conception of the model reported here was determined by considering oil prices as well as the extent of foreign aid and of similar transfers of wealth.

Nontrivial structural problems arise only with nonlinear models. In this case one may consider the phase space of the variables in question. There are usually singular points, saddle points, sources, or sinks, which imply the existence of basins. These basins are separate and are therefore divided from each other by separatrices (see Appendix A) consisting of one or of infinitely many trajectories. The evolution of the trajectories may be generally desirable with regard to one basin but undesirable in others. To demonstrate such features, Häfele (1975) conceived a simple model with population and per capita energy demand as the only variables. It then appeared desirable to consider a more powerful and, one might hope, more meaningful model permitting us, among other possibilities, to discuss questions of capital costs for new energy technologies.

This interest of IIASA's Energy Group coincided with another line of interest at IIASA: Holling and his team were studying the dynamics of ecological systems. One motivation for their study was to develop pest management strategies – an effective spraying policy, for example – in an ecosystem. In the framework of this research, Holling (1973, ed. 1978) used the term *resilience* to describe a system's capability to continue its evolution in the same basin when impacts on the system occur from outside. An IIASA workshop (Grümm ed. 1975) brought together ecologists, economists, and climatologists under Koopman's chairmanship, and Grümm (1976) generalized and formalized the notion of resilience. While the precise mathematical definition of the term is still subject to debate, the concept is obviously helpful and enlightening. Indeed, when considering the problems of the next half century, we are less concerned about quantitative evolutions – which nevertheless shape the overall structure – than about the possibility that the system might collapse and the trajectory continue in a different basin.

The model presented here should be seen in light of these considerations. It was designed, not to provide a final answer to the problem of structural stability while simultaneously dealing with all economic details, but rather to make sense economically and technically. We have proposed a sequence of steps to reach the final goal. So that this methodological development is most fruitful, we hope that others will help to improve the state of the art. Broadening our understanding of resilience and including models other than the ecological one would also be desirable. The modeling effort reported here is clearly experimental.

2 RESILIENCE

At present there are several slightly different concepts of resilience, all of which stem from Holling's work (1973). We shall briefly describe the concept preferred by IIASA's Energy Systems Program. It is strongly tied to the theory of differentiable dynamical systems, i.e., the global theory of differentiable equations (see Appendix A); the mathematical definitions of resilience (given in Grümm 1976) are expressed in terms of that theory.

Resilience – conceptually described – is the ability of systems to withstand exogenous, incontrollable disturbances affecting the values of state variables and parameters without qualitatively changing their behavior. As this is originally a property of the system existing in reality, resilience is reflected in the mathematical model describing that system. As we tend to identify the system with the model, we also speak of resilience as a property of the latter. Models used in this context describe the evolution of the system as motion in a state space each of whose points uniquely identifies a state of the system. We assume that the motion is given by a causal (as opposed to a stochastic or nonautonomous) differential equation on state space, thus bringing the results of dynamical systems theory to bear. For any given set of parameter values, the state space contains one or several attractors that describe steady modes of the system's behavior. The basins of these attractors are separated by basin boundaries. All parameter values corresponding to the same structure of state space form one parameter regime (economic niche). Resilience may be described as follows: if, due to the change in state variables/parameters, the state/set of parameter values has not left its previous basin/parameter regime, the system has absorbed the disturbance; if it has jumped over a boundary, qualitative – often catastrophic – changes will occur.

As basins and parameter regimes are usually open sets, disturbances below a certain level will always be absorbed, thus resolving at the same time our uncertainty about fine details of the dynamics, as well as about the exact values of state variables and parameters. Small differences between the exact values and the values taken for the mathematical model will not qualitatively affect the model's output.* This form of the "structural stability" argument is, of course, well known from catastrophe theory.

Numerical measures for resilience can be introduced next, many of them incorporating the notion of "distance to some boundary." Full details are contained in *Definitions of Resilience* (Grümm 1976).

In our model, resilience with respect to changes in state variables is formally trivial: there is at most one basin. However, the constraint that small disturbances or errors in the state variables should not lead to drastically different future behavior translates into the postulate – as shall be seen – that the system has an attractor, which corresponds to an economically meaningful state of the model economy. As there are different parameter regimes (niches), resilience with respect to parameter changes is a meaningful question. This leads to the problem of determining the boundaries of these regimes in parameter space; if the parameters of an "actual" economy come close to a boundary, one should begin to be concerned.

3 HISTORICAL DEVELOPMENT

In this section we shall summarize the evolution of highly aggregated economic models at IIASA in the past three years. We shall give only brief descriptions of four models (A to D), as full details are contained in *Economy Phase Portraits* (Grümm and Schrattenholzer 1976).

Model A (Häfele 1975) showed for the first time in an economic model at IIASA how a saddle point, by generating a separatrix, can cause two basins with different long-term behavior to occur. The model attempts to describe phenomenologically different effects: the influence of the standard of living on the birth rate and on the level of safety expenditures, as well as the rise of energy consumption with an increase in the gross national product (GNP). The model has population and per capita energy consumption as state variables; its two basins describe two trends: toward "low population living in luxury" on the one hand and toward "growing population at a constant living standard" on the other.

Model B (Avenhaus, Grümm, and Häfele 1975) is closer to established economic formulation. The entire economy is split into two parts, with an energy-producing sector distinguished from the rest. The other assumptions of Model A are incorporated, although

*We can view these differences as "exogenous, incontrollable perturbations"!

their mathematical expression is necessarily different. The state variables are population, GNP per capita, and energy production; total GNP is split into consumption, depreciation, and net investment. A ceiling is assumed for per capita GNP. The two attractors and basins of the model correspond to two possibilities for producing this limiting GNP, one with a high energy production and a low investment in the nonenergy sector and the other with the reverse situation.

Model C is the result of combining Model B with ideas presented in Häfele and Bürk (1976). Labor is introduced in addition to energy and nonenergy capital stock as a production factor. A new feature is that the dynamics of this model are given by an infinitesimal optimization postulate: at each point in state space, we should proceed in the direction that optimizes the rise of GNP. Various phase portraits for this model are given by Grümm and Schrattenholzer (1976).

Total GNP has thus far been given by a Cobb–Douglas ansatz for the production function. In Model D the investment sector is described by a linear input–output ansatz. The fraction of total GNP available for investments (denoted by $1 - \alpha$) plays a central role and is determined dynamically. Depending on the parameter values, the system has two to four basins, but its attractors are not isolated fixed points: along entire curves the system is at equilibrium.

As the following description makes clear, many of the "building blocks" of the present model are contained in these four models. The overall structure of our model owes much to the work of Häfele and Bürk (1976).

Two further, unpublished "mini-models" by Bürk and Grümm give a phenomenological treatment of the price rise for a scarce resource and thereby justify the "logistic transition between two different technologies" assumed in the present model.

4 THE MODEL'S STRUCTURE

We selected energy, capital, and labor force, denoted by E, K, and L, respectively, as basic variables spanning the state space of our model. Their precise economic relevance is demonstrated in Fig. 1. E, K, and L denote the respective stocks of

- Total installed power (we also refer to this as the total invested energy-related capital stock)
- Total invested non-energy-related capital stock
- Total available stock of skilled labor

Illustrations of these quantities follow. (In the phenomenological spirit of the model, we do not give economically exact definitions.)

E includes, for example, power stations, with their integrated equipment, such as turbines, generators, dams (in the case of hydropower), electric grids, and so forth; oil refineries; pipelines; and tankers. As we assume a constant load factor, E may also be interpreted as the total energy output (or input into the economy) per time unit. In this report we use 1 year as the time unit and 0.75 as the load factor. Hence 1 W of installed power yields 0.75 Wyr of energy per year.

K denotes all factories in operation (assuming no spare capacity), with their machinery and equipment to produce

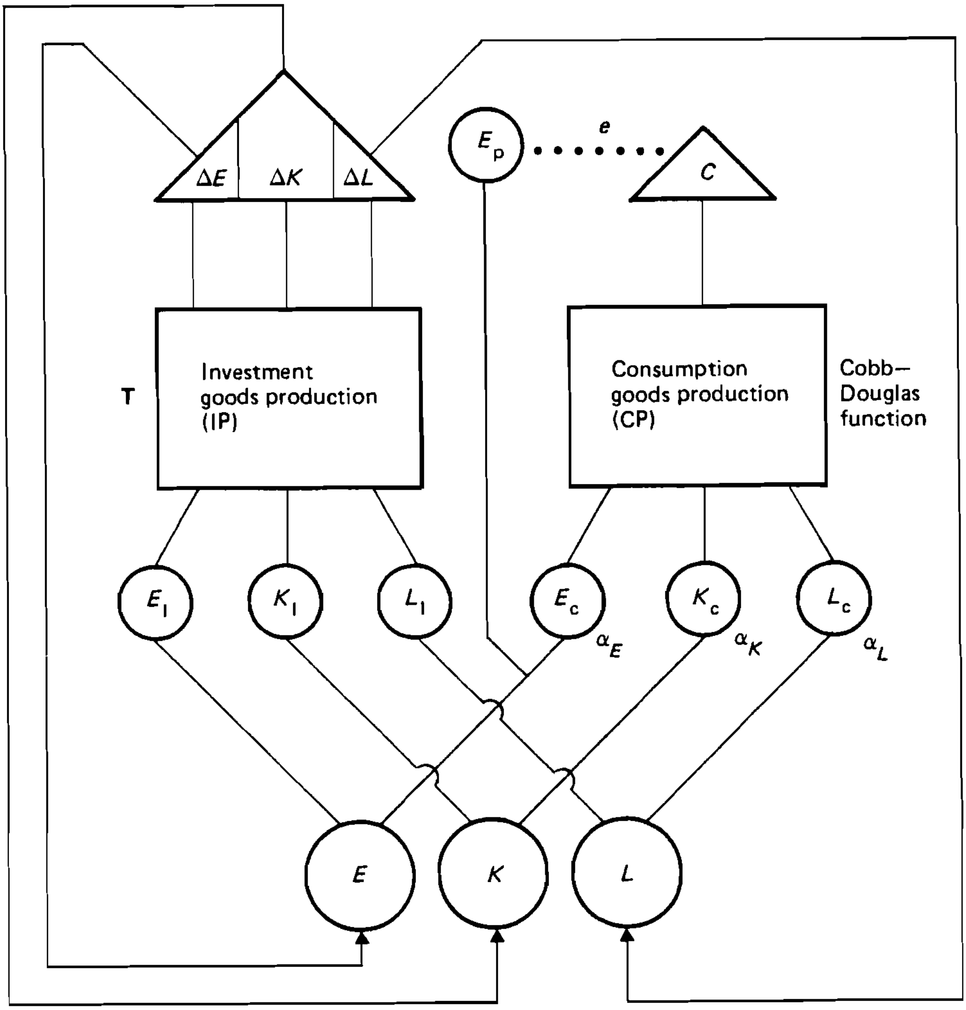


FIGURE 1 Structure of the model.

- Energy-related capital goods (the stock denoted by E), such as turbines; generators; and cement for existing dams and power stations
- Non-energy-related capital goods, such as new machinery (which may eventually produce new turbines) or cement for new factories, schools, universities, or other means of "skill production" (but not cement for new dams, power stations, or private homes)
- Consumption goods, such as private cars, private homes (i.e., cement to build private homes), and everything else that does not produce anything in turn

K also denotes existing schools, universities, and other means of "skill production." The unit for K is \$1.

L represents the available stock of labor (assuming no idle workers) weighted with skill: effectiveness, know-how, sophistication of such tools as pocket calculators, and so forth.* With a load factor of 0.25, 1 worker of unit skill corresponds to 2000 effective man-hours per year or 0.25 effective man-years per year. By introducing skill, we are able to increase L without increasing the number of workers and to omit a term $\sim e^{mt}$ (describing technological progress) from our production functions.

We assume two lines of production within our system. The first is investment goods production (IP), which produces (per year)

- Energy-related capital goods ΔE
- Non-energy-related capital goods ΔK
- Skilled labor force ΔL

The second is consumption goods production (CP), which produces consumption goods C (per year). C does not include private energy consumption. Energy is required to utilize and to maintain consumption goods, especially such middle- and long-term durable goods as cars and homes. We call the part of E allocated for this purpose E_p (private energy consumption). Because even homes are of finite durability (say, 50 years), the appropriate portion is in fact "consumed" each year. Hence in a first approximation E_p is assumed to be proportional to C:

$$E_p = eC \tag{1}$$

This assumption is backed by statistics.

Combining what has been said about the stocks and their allocation, we have

$$\begin{aligned} E &= E_I + E_c + E_p \\ K &= K_I + K_c \\ L &= L_I + L_c \end{aligned} \tag{2}$$

We express the stocks allocated to CP in terms of the total quantities, using the coefficients $\alpha_E, \alpha_K, \alpha_L$:

$$E_c = \alpha_E E \qquad K_c = \alpha_K K \qquad K_c = \alpha_L L \tag{3}$$

The αs clearly describe the emphasis on CP within the economy.

The inputs into IP are E_I, K_I, and L_I; the outputs, ΔE, ΔK, and ΔL. Part $(d_x X)$ of the outputs has to compensate for the respective depreciations (in the case of skilled labor, retirements of laborers); the other part $(\dot{X})$ is the annual net increase in the respective stocks:

$$\Delta X = \dot{X} + d_x X \qquad X = E,K,L \tag{4}$$

where the d_x denote the respective depreciation rates.

*Again, we do not define *skill* quantitatively; we assume that it describes both "subjective" (e.g., better training) and "objective" (e.g., better technology) increases in productivity. We could use a similar concept of *effective capital.*

Common sense tells us that stocks and outputs have to be nonnegative, i.e.,

$$E \geqslant 0 \qquad K \geqslant 0 \qquad L \geqslant 0$$
$$\Delta E = \dot{E} + d_E E \geqslant 0 \qquad \Delta K = \dot{K} + d_K K \geqslant 0 \qquad \Delta L = \dot{L} + d_L L \geqslant 0 \tag{5}$$

Thus our state space is R^3_+.

We now assume that the IP is of the linear input–output type with minimum production functions (see Dorfman, Samuelson, and Solow 1958; Samuelson 1947, 1976; Beaumol 1977; and Hicks 1969)

$$\Delta X = \min_{Y=E,K,L} (Y_{\mathrm{I},X}/a_{YX}) \qquad X = E,K,L \tag{6}$$

More explicitly,

$$\Delta E = \min\left(\frac{E_{\mathrm{I},E}}{a_{EE}}, \frac{K_{\mathrm{I},E}}{a_{KE}}, \frac{L_{\mathrm{I},E}}{a_{LE}}\right) = \dot{E} + d_E E \geqslant 0$$

$$\Delta K = \min\left(\frac{E_{\mathrm{I},K}}{a_{EK}}, \frac{K_{\mathrm{I},K}}{a_{KK}}, \frac{L_{\mathrm{I},K}}{a_{LK}}\right) = \dot{K} + d_K K \geqslant 0 \tag{6'}$$

$$\Delta L = \min\left(\frac{E_{\mathrm{I},L}}{a_{EL}}, \frac{K_{\mathrm{I},L}}{a_{KL}}, \frac{L_{\mathrm{I},L}}{a_{LL}}\right) = \dot{L} + d_L L \geqslant 0$$

a_{YX} denotes the minimal amount of stock Y_I (the portion of the total stock Y allocated to IP) required to produce one unit of output of stock of type X; correspondingly, $Y_{\mathrm{I},X}$ is the total amount of Y_I necessary to produce ΔX. Another way of looking at eq. (6′) is

$$a_{YX}\,\Delta X \leqslant Y_{\mathrm{I},X} \text{ for all } X,\ Y = E,K,L \tag{7}$$

Adding the left- and right-hand sides for $X = E,K,L$ while keeping Y fixed, we obtain

$$\sum_{X=E,K,L} a_{YX}\,\Delta X \leqslant \sum_{X=E,K,L} Y_{\mathrm{I},X} = Y_\mathrm{I} \qquad Y = E,K,L \tag{8}$$

Hence $(Y_{\mathrm{I},E}, Y_{\mathrm{I},K}, Y_{\mathrm{I},L})$ is a particular allocation of Y_I to the three production lines within the IP.

We first assume that our IP runs optimally – that is, without slacks. Thus all ratios $Y_{\mathrm{I},X}/a_{YX}$, X fixed, $Y = E,K,L$ over which the minimum has to be taken in eq. (6′) are the same, and the inequality in eq. (8) becomes an equality. In part of state space, this requirement of optimality is inconsistent with eq. (5).

In matrix notation this reads

$$\mathbf{T}\Delta X = X_\mathrm{I} \tag{9}$$

where

$$\Delta X = \begin{pmatrix} \Delta E \\ \Delta K \\ \Delta L \end{pmatrix} \text{ and } X_\mathrm{I} = \begin{pmatrix} E_\mathrm{I} \\ K_\mathrm{I} \\ L_\mathrm{I} \end{pmatrix}$$

$$\mathbf{T} \doteq \begin{pmatrix} a_{EE} & a_{EK} & a_{EL} \\ a_{KE} & a_{KK} & a_{KL} \\ a_{LE} & a_{LK} & a_{LL} \end{pmatrix} \tag{10}$$

is called the technological matrix because it reflects the technological situation within our model economy.

The examples that follow illustrate the significance of the matrix coefficients. (We estimate the values of the coefficients for our base case in Section 6.)

- a_{EE} is the number of watts allocated to IP in order to increase the installed power E by 1 W per year
- a_{EK} is the number of watts allocated to IP in order to increase the non-energy-related capital stock K by \$1 per year
- a_{EL} is the number of watts allocated to IP in order to increase the number of skilled workers L by 1 person per year
- a_{KE} is the amount of capital invested in IP required to increase E by 1 W per year
- a_{KK} is the amount of capital invested in IP required to increase K by \$1 per year
- a_{KL} is the amount of capital invested in IP required to increase L by 1 person per year
- a_{LE} is the number of skilled workers employed in IP required to increase E by 1 W per year
- a_{LK} is the number of skilled workers employed in IP required to increase K by \$1 per year
- a_{LL} is the number of skilled workers employed in IP required to increase L by 1 person per year

The choice of eq. (6) as the production function for the IP implies nonsubstitutability among the production factors. This property seems to be realistic; the lines of production within heavy industry, for instance, are rather inflexible and allow only minor deviations from an optimal path.

In contrast, the production function of the CP should allow for substitutability; using a simple ansatz, we describe it by a Cobb–Douglas function:

$$C = AE_\mathrm{c}^\alpha K_\mathrm{c}^\beta L_\mathrm{c}^\gamma \tag{11}$$

$$\alpha + \beta + \gamma = 1 \tag{12}$$

Equation (12) accounts for constant returns to scale. Combining eqs. (1), (2), (3), (4), (9), and (11), we arrive at the system of ordinary differential equations that we sought and that we shall analyze by global techniques:

$$\mathbf{T}(\dot{X} + \mathbf{d}X) = (1 - \alpha)X - X\mathbf{p} \tag{13}$$

where we have used the notation

$$\mathbf{d} = \begin{pmatrix} d_E & 0 & 0 \\ 0 & d_K & 0 \\ 0 & 0 & d_L \end{pmatrix} \quad \mathbf{1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \alpha = \begin{pmatrix} \alpha_E & 0 & 0 \\ 0 & \alpha_K & 0 \\ 0 & 0 & \alpha_L \end{pmatrix} \quad X_\mathrm{p} = \begin{pmatrix} E_\mathrm{p} \\ 0 \\ 0 \end{pmatrix} \tag{13'}$$

with $E_\mathrm{p} = eA\alpha_E^\alpha \alpha_K^\beta \alpha_L^\gamma E^\alpha K^\beta L^\gamma$.

Equation (13) may be rewritten as

$$\dot{X} = \mathbf{T}^{-1}\{(\mathbf{1} - \alpha - \mathbf{Td})X - X_\mathrm{p}\} \tag{14}$$

Inserting eq. (14) in eq. (5) leads to a condition defining a region in the state space where the optimality assumption is consistent with the positivity requirement, eq. (5). We call this region the slack-free region: inside it, the dynamics of our system (its evolution in time) are given by eq. (14).

Our considerations will be largely restricted to the slack-free region because such questions as the existence of equilibrium states and their stability can be discussed within it. Possible dynamics outside the slack-free region are described in Appendix B; they differ from the formal continuation of eq. (14) to the outside. As shown in the appendix, this difference does not affect the results presented in the following sections, which use eq. (14) in the entire state space.

To analyze eq. (14) we would normally calculate the critical element – in our case, the fixed point (FP). Obviously, $X = 0$ is an FP and we can easily show that, in general, it is the only FP.

We put $\dot{X} = 0$; hence, from eq. (14),

$$(\mathbf{1} - \alpha - \mathbf{Td})X = X_\mathrm{p} \tag{15}$$

From the expansion of eq. (15), we express E and K as linear functions of L and substitute them into the first line. Since E_p, the first component of X_p, shows constant return to scale, the right-hand side of eq. (15) is linear in L; so, also, is the left-hand side. Hence $L \neq 0$ drops out, and we are left with a restriction on the parameters; we cannot expect this restriction to hold generally.

If we assume variable returns to scale in the Cobb–Douglas function (i.e., if instead of eq. (12) we have $\alpha + \beta + \gamma \neq 1$), there will be a second FP in the E,K,L space. We then look for a fixed ray instead of an FP in state space; this ray describes an equilibrium state of stable growth.

The property of constant returns to scale of both production functions allows us to introduce new variables and to reduce three dimensions to two. Inserting

$$\epsilon := E/L \qquad \kappa := K/L \qquad \lambda := \log L \tag{16}$$

in eq. (13′) yields, after straightforward algebra,

$$\begin{pmatrix} \epsilon \\ \kappa \\ \lambda \end{pmatrix}^{\cdot} = \begin{pmatrix} 1 & 0 & -\epsilon \\ 0 & 1 & -\kappa \\ 0 & 0 & 1 \end{pmatrix} \mathbf{T}^{-1} \left\{ (\mathbf{1} - \alpha - \mathbf{Td}) \begin{pmatrix} \epsilon \\ \kappa \\ 1 \end{pmatrix} - \begin{pmatrix} B\,\epsilon^\alpha\,\kappa^\beta \\ 0 \\ 0 \end{pmatrix} \right\} = \begin{pmatrix} f(\epsilon,\kappa) \\ g(\epsilon,\kappa) \\ h(\epsilon,\kappa) \end{pmatrix} \tag{17}$$

where $B = eA\alpha_E^{\alpha}\alpha_K^{\beta}\alpha_L^{\gamma}$. Thus $\dot{\lambda}$ is the growth rate of skilled labor. We observe that the right-hand side of eq. (17) depends only on ϵ and κ – not on λ; the problem has become two-dimensional. One first solves for ϵ and κ and then substitutes the solutions into the expression for $\dot{\lambda}$ and integrates. Details are given in Section 5. The system parameters, which must be chosen consistently, are as follows: nine coefficients of **T**, the technological matrix, $a_{XY}, X, Y = E, K, L$; three ratios, $\alpha_E, \alpha_K, \alpha_L$; three depreciation rates, d_E, d_K, d_L; and three Cobb–Douglas coefficients, α, β, and B (where A and e are combined). These parameters make up a parameter space of eighteen dimensions.

5 ANALYSIS OF THE MODEL EQUATIONS

For a qualitative analysis of eq. (17), we follow the standard procedure:

- Determination of FPs and their properties
- Determination of the slack-free region according to eq. (5)
- In the case of a stable FP, determination of the central region, i.e., the region of initial values within the slack-free region such that the trajectories do not subsequently leave the slack-free region

As previously mentioned, an FP $\dot{\epsilon} = \dot{\kappa} = 0$ corresponds to a time-invariant ray (direction) in E,K,L space. Equations (18) and (19) are valid only inside the slack-free region. This amounts to solving the (nonlinear) eigenvalue problem

$$\dot{X} = \mathbf{T}^{-1}\{(1 - \alpha - \mathbf{Td})X - X_p\} = nX \tag{18}$$

in terms of eq. (14) or

$$\mathbf{T}^{-1}\left\{(1 - \alpha - \mathbf{Td})\begin{pmatrix}\epsilon\\ \kappa\\ 1\end{pmatrix} - \begin{pmatrix}B\ \epsilon^{\alpha}\ \kappa^{\beta}\\ 0\\ 0\end{pmatrix}\right\} = n\begin{pmatrix}\epsilon\\ \kappa\\ 1\end{pmatrix} \tag{19}$$

in terms of eq. (17).

To get a feeling for the situation, we may analyze a simplified version of eq. (18) or (19). We use the argument of structural stability and extrapolate eqs. (18) and (19) outside the slack-free region. We assume all depreciations to be equal, $d_E = d_K = d_L = d$, consider X_p as a perturbation that we put to zero in our simplification,* and solve the remaining linear problem. n then appears as the simultaneous growth rate of E, K, and L at the point of stable growth, i.e., at the fixed ray.

Simple algebra then leads to the respective equations

$$[(1 - \alpha)^{-1}\mathbf{T} - z]\,X = 0 \tag{18'}$$

or

$$[(1 - \alpha)^{-1}\mathbf{T} - z]\begin{pmatrix}\epsilon\\ \kappa\\ 1\end{pmatrix} = 0 \tag{19'}$$

where $z = 1/(d + n)$.**

---

*This is justified by looking at actual numerical values.

**z is an auxiliary quantity introduced to apply a Frobenius theorem.

As $(1 - \alpha)^{-1}\mathbf{T}$ is a matrix with strictly positive elements, a theorem of Frobenius is applicable. This theorem implies that exactly one eigenvector with positive components exists and that its corresponding eigenvalue is positive and is the largest among the real eigenvalues. As eqs. (18′) and (19′) are of third degree, there are two possibilities: either three real solutions for z (and hence for n) or one real and two complex conjugate solutions.

In the first case, we have three fixed points in the ϵ–κ plane with the respective

$$n_i = 1/z_i - d \qquad i = 1,2,3 \tag{19''}$$

as growth rates. The FP corresponding to the smallest growth rate is in the positive quadrant and is a sink. The other two are a saddle point and a source, respectively, but their location is outside the positive quadrant. Qualitatively, the full phase portrait then appears as shown in Fig. 2.

In the second case, we have only one FP, with a growth rate $n > -d$. It is again located within the positive quadrant and may be attractive or repulsive; the corresponding phase portraits are shown in Fig. 3.

Owing to structural stability, these statements also remain true, within certain limits, for differing ds and nonvanishing X_p. For the base cases that we studied, the second alternative holds; it prevails even for large deviations from the base case data. Appendix C gives a criterion for distinguishing between the two alternatives.

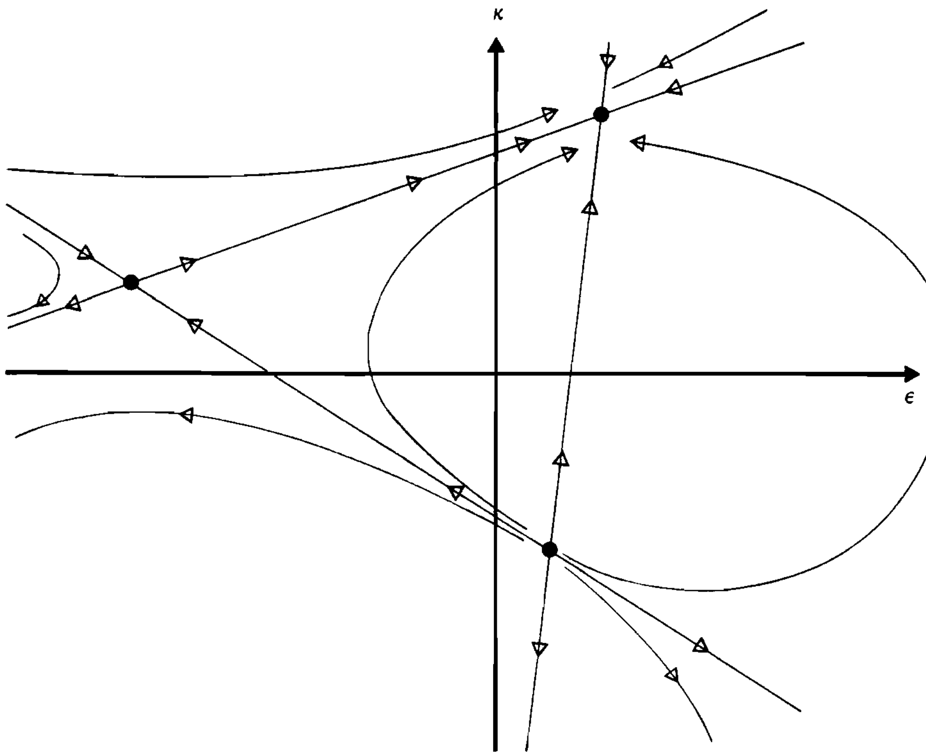


FIGURE 2 Three real fixed points.

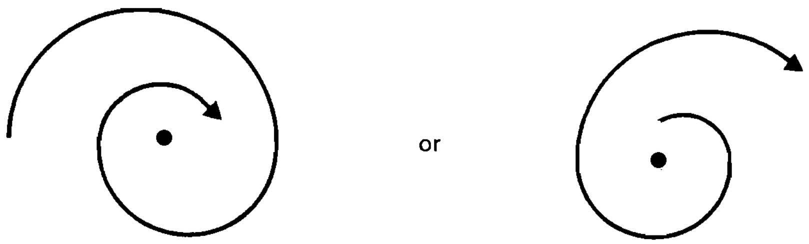


FIGURE 3 One real fixed point.

Returning to eq. (19), the fixed point condition $\dot{\epsilon} = \dot{\kappa} = 0$ yields (ϵ_0, κ_0) as a solution of

$$f(\epsilon,\kappa) = 0 \text{ and } g(\epsilon,\kappa) = 0 \tag{20}$$

Subsequent substitution into h leads to

$$n_0 = h(\epsilon_0, \kappa_0) \tag{21}$$

where the functions f, g, and h are defined by eq. (17). This is the simultaneous growth rate of E, K, and L at the FP.

For purely technical reasons it is advantageous first to solve

$$g(\epsilon,\kappa) = 0 \text{ and } h(\epsilon,\kappa) = n \tag{22}$$

to obtain ϵ and κ as functions of n, and then to substitute them into $f(\epsilon,\kappa) = 0$ and to solve for n. Thus, after some manipulation,

$$\begin{aligned}
&(d_E + n)a_{KE}\epsilon + [(d_K + n)a_{KK} - (1 - \alpha_K)]\,\kappa + (d_L + n)a_{KL} = 0 \\
&(d_E + n)a_{LE}\epsilon + (d_K + n)a_{LK}\kappa + (d_L + n)a_{LL} - (1 - \alpha_L) = 0 \\
&[(d_E + n)a_{EE} - (1 - \alpha_E)]\,\epsilon + (d_K + n)a_{EK}\kappa + B\epsilon^{\alpha}\kappa^{\beta} + (d_L + n)a_{EL} = 0
\end{aligned} \tag{23}$$

These equations must be solved for ϵ, κ, and n, respectively. With the abbreviations

$$\begin{aligned}
&D_1 := a_{KE}a_{LK} - a_{KK}a_{LE} \qquad D_2 := a_{KK}a_{LL} - a_{KL}a_{LK} \\
&D_3 := a_{KE}a_{LL} - a_{KL}a_{LE} \qquad D(n) := (d_E + n)[a_{LE}(1 - \alpha_K) + (d_K + n)D_1]
\end{aligned} \tag{24}$$

we obtain

$$\epsilon(n) = \frac{1}{D(n)}\,[(1 - \alpha_K)(1 - \alpha_L) - (1 - \alpha_K)(d_L + n)a_{LL} - (1 - \alpha_L)(d_K + n)a_{KK}$$

$$+ (d_K + n)(d_L + n)D_2]$$

$$\kappa(n) = \frac{d_E + n}{D(n)} [(1 - \alpha_L)a_{KE} - (d_L + n)D_3] = \frac{(1 - \alpha_L)a_{KE} - (d_L + n)D_3}{(1 - \alpha_K)a_{LE} + (d_K + n)D_1} \tag{25}$$

Substitution of eq. (25) into the third part of eq. (23) yields a transcendental equation for n.

Three remarks are in order. First, as previously mentioned, the simplified system of equations (all ds equal, $X_p = 0$) always has a solution with $\epsilon \geqslant 0, \kappa \geqslant 0$. On the other hand, the general set of equations G need not have a solution at all; there need not exist a domain for n where both $\epsilon(n)$ and $\kappa(n)$ are nonnegative (they obviously should be, as eq. (5) demonstrates).

Second, n always appears in connection with depreciation rates; hence, changing the three depreciation rates by the same amount and simultaneously changing n by the same amount (but with the opposite sign) leaves ϵ_0 and κ_0 unchanged.

Third, as functions $\epsilon(n)$ and $\kappa(n)$ depend on the 18 system parameters, the previously mentioned positivity condition restricts the allowed parameter values to certain regions or economic niches within the parameter space.

To analyze the properties of the FP (ϵ_0, κ_0) (assuming the conditions of eq. (5) to be fulfilled), we follow the standard procedure: linearization of eq. (17) yields, after straightforward calculations,

$$\begin{aligned} \begin{pmatrix} \epsilon' \\ \kappa' \\ \lambda' \end{pmatrix}^{\bullet} &= \begin{pmatrix} 0 \\ 0 \\ n_0 \end{pmatrix} \\ &+ \left\{ \begin{pmatrix} 1 & 0 & -\epsilon_0 \\ 0 & 1 & -\kappa_0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{T}^{-1} \left[\mathbf{1} - \alpha - \mathbf{Td} - \begin{pmatrix} \alpha B \epsilon_0^{\alpha-1} & \kappa_0^{\beta} & \beta B \epsilon_0^{\alpha} & \kappa_0^{\beta-1} & 0 \\ & 0 & & 0 & 0 \\ & 0 & & 0 & 0 \end{pmatrix} \right] - n_0 \mathbf{1} \right\} \begin{pmatrix} \epsilon' \\ \kappa' \\ 0 \end{pmatrix} \\ &+ \ldots = \begin{pmatrix} 0 \\ 0 \\ n_0 \end{pmatrix} + \mathbf{S} \begin{pmatrix} \epsilon' \\ \kappa' \\ 0 \end{pmatrix} + \ldots \end{aligned} \tag{26}$$

where we introduced

$$\epsilon' = \epsilon - \epsilon_0 \qquad \kappa' = \kappa - \kappa_0 \qquad \lambda' = \lambda - \lambda_0 \tag{26'}$$

and where

$$\mathbf{S} = \begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{pmatrix}$$

denotes the matrix between the braces.* Considering the first two parts of eq. (22), we have, with

$$\mathbf{S}_1 := \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

*Explicit expressions for matrix elements of S tend to become cumbersome; as they must be evaluated numerically in any case, they are omitted here.

$$\begin{pmatrix}\epsilon' \\ \kappa'\end{pmatrix}^{\bullet} = S_1 \begin{pmatrix}\epsilon' \\ \kappa'\end{pmatrix} + \ldots \tag{27}$$

According to the general theory, the behavior of the solutions near the FP is determined by the eigenvalues $\mu_{1,2}$ of $\mathbf{S}_1$, which are given by

$$\mu_{1,2} = 1/2\{\text{tr}\,\mathbf{S}_1 \pm [(\text{tr}\,\mathbf{S}_1)^2 - 4\,\det\,\mathbf{S}_1]^{1/2}\}^* \tag{28}$$

The FP (ϵ_0, κ_0) is stable if the real parts of $\mu_{1,2}$ are negative

$$\text{Re}\,\mu_{1,2} < 0 \tag{29}$$

It is called real stable if $\text{Im}\,\mu_{1,2} = 0$ and complex stable in any other case. For both real stable and complex stable FPs, the trajectories $(\epsilon(t), \kappa(t))$ approach the FP in the future. If eq. (29) does not hold, tbe FP is unstable; with the exception only of special cases, the trajectories leave any neighborhood of the FP in the future, even if there are finite time periods during which the FP is approached.

Analyzing eq. (28) in more detail, we may have

$$(\text{tr}\,\mathbf{S}_1)^2 \geqslant 4\,\det\,\mathbf{S}_1 \tag{30}$$

or

$$(\text{tr}\,\mathbf{S}_1)^2 < 4\,\det\,\mathbf{S}_1 \tag{30$'$}$$

In the case of eq. (30′), the $\mu_{1,2}$ will be real. We now have to distinguish

$$\det\,\mathbf{S}_1 > 0 \tag{31}$$

from

$$\det\,\mathbf{S}_1 < 0 \tag{31$'$}$$

for eq. (31). Both μs have the same sign as tr $\mathbf{S}_1$. Hence

$$\det\,\mathbf{S}_1 > 0 \text{ and } \text{tr}\,\mathbf{S}_1 < 0 \tag{32}$$

is the stable case;

$$\det\,\mathbf{S}_1 > 0 \text{ and } \text{tr}\,\mathbf{S}_1 > 0 \tag{32$'$}$$

is the unstable case.

For eq. (31′) the μs have the opposite sign and the FP is unstable. In the case of eq. (30′), the eigenvalues will be conjugate complex, with

$$\text{Re}\,\mu_1 = \text{Re}\,\mu_2 = \frac{1}{2}\text{tr}\,\mathbf{S}_1 \tag{33}$$

Hence tr $\mathbf{S}_1 > 0$ means instability, and tr $\mathbf{S}_1 < 0$, stability, of the FP.

---

*Here tr $\mathbf{S}_1 = S_{11} + S_{22} = \mu_1 + \mu_2$ (11′) denotes the trace of S, and det $\mathbf{S}_1 := S_{11}S_{22} - S_{12}S_{21} = \mu_1 \cdot \mu_2$ (11″) denotes the determinant of $\mathbf{S}_1$, respectively.

Combining these results, we can say that eqs. (30), (31), and (32) yield a real stable FP, whereas eqs. (30′) and (32) yield a complex stable FP. All other situations are unstable.

The economic relevance of the stability or instability of an FP is significant. The ratios $\epsilon = E/L$ and $\kappa = K/L$ attain certain constant values $\epsilon_0 = E_0/L_0$ and $\kappa_0 = K_0/L_0$, respectively, at the FP. This means that E, K, and L have the same time evolution, given by n_0 of eq. (21), at the point (E_0,K_0,L_0).

If the FP (ϵ_0,κ_0) is unstable, trajectories will move away from it in time, and the system will move into unrealistic regions, i.e., arbitrarily large or small values of E/L and K/L. Although our model is unrealistic for such values of the state variables, we can interpret this behavior as a prediction of catastrophic, and certainly undesirable, behavior of our model economy. If, on the other hand, the FP is stable, the system will tend to stable values of E/L and K/L and will achieve stable growth (or decrease, if n_0 happens to be negative). Although at this point we can draw this conclusion only for evolution within the feasible region, so that eq. (18) is valid, Appendix B shows it also holds in a neighborhood of the slack-free region. This is certainly a more desirable economic situation.

In the case of stability, we can distinguish within the slack-free region a central region consisting of points the entire evolution of which will remain in the slack-free region (e.g., A in Fig. 4). In contrast, points such as B in Fig. 4 will for some time leave the slack-free region, although they, too, will come back and tend toward the FP. Thus their evolution will be governed for some time by the dynamics discussed in Appendix B, which imply large-scale variations of the economy. We may regard the central region as the set of "best initial conditions" because a smooth evolution toward stable growth is assured there.

We note that the central region is different from the slack-free region only if the FP is a stable focus (i.e., has two complex eigenvalues with negative real parts). For an unstable FP, the concept of a central region is meaningless, as all trajectories except that coinciding with the FP will leave the slack-free region.

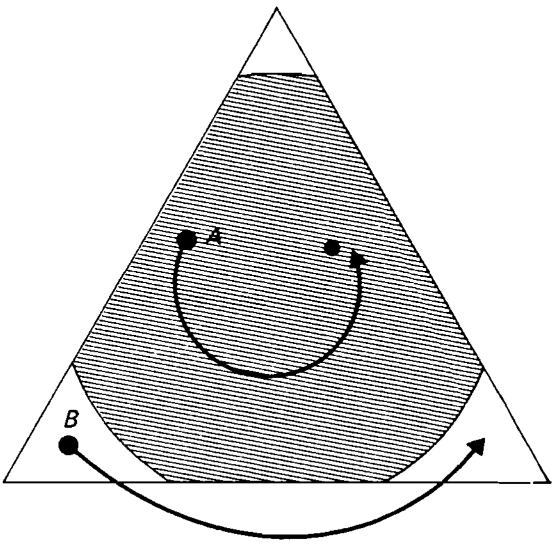


FIGURE 4 The slack-free region.

Note, too, that eq. (18) becomes undefined if **T** is not invertible, i.e., if det **T** = 0. For such a technology matrix, E_I, K_I, L_I would be restricted to a plane. We assume that this is not the case, from a genericity argument. Care must be taken, however, if | det **T** | becomes too small or if, by an adiabatic variation of the technology matrix (see Section 6), we should cross the hypersurface in parameter space, where det **T** = 0.

6 THE BASE CASE

We chose a base set of parameters for actual calculations. As the elements of the technology matrix **T** describe the technological standard of the model economy, the numbers could be expected to differ significantly according to their correspondence to the situation of a developed country (DC) or to that of a less developed country (LDC). In the case of a DC, most could be taken directly from data available at IIASA, while some had to be deduced from statistical material. In particular, the last column of **T**, referring to skill production, required comparison of the relative numbers of teachers and students, identification of the depreciation rate with the rate of retirements, and so forth. In the case of an LDC, orders of magnitude of the required numbers were obtained from LDC specialists; these estimates are necessarily crude.

The primary purpose of the numerical calculations was not to obtain "predictions" but rather to acquire some feeling for the position and size of the economic niches. More explicitly, Section 5 contained the first step toward a division of the parameter space into economically meaningful parts and useless regions, according to the character of the FP. The second step involves selection of definite base case values within an economically meaningful part and estimation of the size of the region around these values, such that the qualitative behavior of the base case remains unchanged. For a stable FP we call this region a favorable economic niche. Favorable and unfavorable economic niches are separated by hypersurfaces; it is interesting to consider which parameters are primarily responsible for crossing such boundaries. In other words, which parameters allow the least range of variation? Since stability or instability of the FP (i.e., favorableness or unfavorableness of the corresponding economic niche) is described by the eigenvalues λ_i (eq. (26)), this amounts to analyzing which parameters show the strongest influence on the λ_i.

The eigenvalues are, however, not the only indicators of the economy; the growth rate at the FP is also important. Even within a favorable niche, the growth rate may be negative, which means that a trajectory will move toward equilibrium, but with shrinking E, K, and L. As previously mentioned, growth rate n always appears in connection with depreciations. The unwanted situation of a negative growth rate could therefore immediately be improved by reducing the depreciations, i.e., by producing goods of higher quality and greater durability.

The model was implemented on a desk computer (HP8830A) and on IIASA's PDP11/70. Implementation allows

- Alternation between countries (i.e., parameter sets (DC and LDC)) and variation of scenario variables
- Computation of the FP and its eigenvalues
- Plotting of the feasible region

- Numeric integration and plotting of specific system trajectories
- Adiabatic variations of scenario variables during the numeric integration

With regard to the last point, the model implementation should be able to simulate an "economy in transition" in the sense used by Häfele *et al.* (1976) – for example, an economy changing from conventional to nuclear energy production. One must thus change some parameters continuously while running the model. We have assumed a transition from initial to final values using a logistic curve; this assumption is confirmed by the data collected by Marchetti and Nakicenovic (1979). The term *adiabatic* refers to the system's smooth response to parameter changes if the time scale of those changes is long compared to that of the system.

Table 1 shows the parameter values of the base case representing a DC, together with the band width of allowed variation of each parameter (all other parameters remaining fixed). Numbers with a single asterisk do not denote the boundary of the niche but rather the point of transition to a negative growth rate. Numbers followed by $<$ are still within the favorable niche, but we did not pursue the upper limit further.

The corresponding FP is located at ϵ_0 = 10 600 Wyr/smyr, κ_0 = 16 700 \$yr/smyr, with a growth rate of $n_0 = 3.3$ percent and eigenvalues $\lambda = -0.047 \pm 0.31i$.

The extent of the favorable niche in some parameter directions is quite wide, while in others (particularly a_{KK} and a_{LL}) it is quite narrow.

Table 2 shows a set of parameter values that might represent the economy of some LDC. In this case the FP is not stable and we again indicate the extension of the unfavorable niche in each direction.

The corresponding FP is located at ϵ_0 = 7 500 Wyr/smyr, κ_0 = 360 \$yr/smyr, with a growth rate of $n_0 = 2.2$ percent and eigenvalues $\lambda = +0.58 \pm 0.66i$.

Obviously, the extent of the niches in each direction changes if the parameters are not kept fixed at their base case values.

We may think of the base cases as the representation of certain scenarios: **T** describes a particular standard of investment goods production, e describes how effectively the energy allocated to private consumption is used, the αs describe the emphasis on CP within the economy, and the a_{ij} describe a particular technology.

In order to study the effect of a scenario change, we introduced scenario variables, which enable one to study the economic niches in parameter space without needing to vary 18 parameters at the same time; moreover, most independent variations of parameters are unrealistic. Each set of values of the scenario variables, however, is assumed to represent at least a consistent model economy.

H_1 accounts for a transition from standard to new (e.g., nuclear or solar) energy options. We model the full transition by increasing a_{EE} by a factor of 30 and a_{KE} by a factor of 10, i.e.,

$$\begin{aligned} a_{EE} &\to a_{EE}(1 + 29\,H_1) \\ a_{KE} &\to a_{KE}(1 + 9\,H_1) \qquad 0 \leqslant H_1 \leqslant 1^* \end{aligned} \tag{34}$$

Intermediate stages are represented by intermediate values of H_1; this is similar for the other scenario variables.

$^*H_1$ may also be taken as larger than one, which corresponds to still more expensive energy options.

TABLE 1 Parameter values of the base case (DC).

Technological matrix			
a_{EE} = 0.04 yr (0–0.7)	a_{EK} = 3.4 Wyr/\$ (2–100 <)	a_{EL} = 300 Wyr/smyr** (0–22 400)	
a_{KE} = 0.1 \$yr/W (0–1.9)	a_{KK} = 2.5 yr (0–3)	$a_{KL} = 1.8 \times 10^4$ \$yr/smyr (1.4×10^4–2×10^5<)	
$a_{LE} = 8 \times 10^{-5}$ smyr yr/W (2.5×10^{-5} -2×10^{-3}* – 3×10^{-3})	$a_{LK} = 8 \times 10^{-5}$ smyr yr/\$ (4×10^{-5}–1.5×10^{-3}*)	a_{LL} = 0.7 yr (0–1)	
Depreciation rates			
d_E = 0.04	d_K = 0.04	d_L = 0.025	
*Consumption fractions****			
α_E = 0.35 (0–0.4)	α_K = 0.75 (0–0.85)	α_L = 0.8 (0–0.9)	e = 0.9 W/\$ (< 0–1.8)
Cobb–Douglas constants			
α = 0.12 (< 0–0.17)	β = 0.12 (< 0–0.17)	A = 425	

*Denotes point of transition to a negative growth rate.
**Skilled man-year.
***αs are varied simultaneously.

TABLE 2 Parameters of the base case (LDC).

Technological matrix			
a_{EE} = 0.04 yr (0–1 <)	a_{EK} = 0.8 Wyr/$ (0.03–7)	a_{EL} = 40 Wyr/smyr* (0–160)	
a_{KE} = 0.02 $yr/W (0–0.2 <)	a_{KK} = 3 yr (2.4–8 <)	a_{KL} = 500 $yr/smyr (0–700)	
$a_{LE} = 2 \times 10^{-4}$ smyr yr/W ($0–3.3 \times 10^{-4}$)	$a_{LK} = 1.4 \times 10^{-3}$ smyr yr/$ ($0–2.4 \times 10^{-3}$)	a_{LL} = 0.4 yr (0.13–4)	
Depreciation rates			
$d_E = 0.02$	$d_K = 0.02$	$d_L = 0.02$	
*Consumption fractions***			
$\alpha_E = 0.8$ (0–0.83)	$\alpha_K = 0.8$ (0–0.83)	$\alpha_L = 0.9$ (0–0.94)	e = 0.3 W/$ (0–0.43)
Cobb–Douglas constants			
$\alpha = 0.4$ (0–0.42)	$\beta = 0.4$ (0–0.42)	$A = 16$	

*Skilled man-year.
**αs are varied simultaneously.

Looking at Table 1, we might expect $H_1 = 1$ to bring the system close to the boundary of the base case niche, if not beyond. Our interpretation – always within the limits of the model – might then be that the base case economy could hardly afford to replace conventional energy options totally by new technology without paying the price somewhere else – on the consumption side, for example. H_2, the second scenario variable, describes the emphasis on CP. We assume, for the sake of simplicity, that the αs are changed simultaneously.

$$\alpha_{E,K,L} \rightarrow \alpha_{E,K,L}(1 + 0.2H_2) \qquad -1 \leqslant H_2 \leqslant 1 \tag{35}$$

$H_2 = -1$ represents a 20 percent reduction of CP inputs; $H_2 = +1$, a 20 percent increase. Not surprisingly, calculations show a high sensitivity of the growth rate n on H_2.

We also examined the effect of increasingly efficient energy use. This would decrease a_{EK} and e but would also change a_{KK}. Classical arguments about substitutability would suggest an increase in a_{KK} ($\sigma = +1$ in the following equations); however, the highly aggregate character of our model makes this argument doubtful. In fact, adherents to the "small is beautiful" school of thought have claimed that saving energy according to that philosophy could actually result in a decrease in a_{KK} ($\sigma = -1$). Thus we have incorporated both alternatives in a scenario variable H_3:

$$\begin{aligned} a_{EK} &\rightarrow a_{EK}(1 - 0.2H_3) \\ a_{KK} &\rightarrow a_{KK}(1 + \sigma 0.2H_3) \qquad \sigma = \pm 1 \\ e &\rightarrow e(1 - 0.3H_3) \qquad 0 \leqslant H_3 \leqslant 1 \end{aligned} \tag{36}$$

Figures 5 and 6 show the boundary in space of scenario variables H_1, H_2, and H_3; this boundary separates a favorable from an unfavorable niche. The direction of scenario

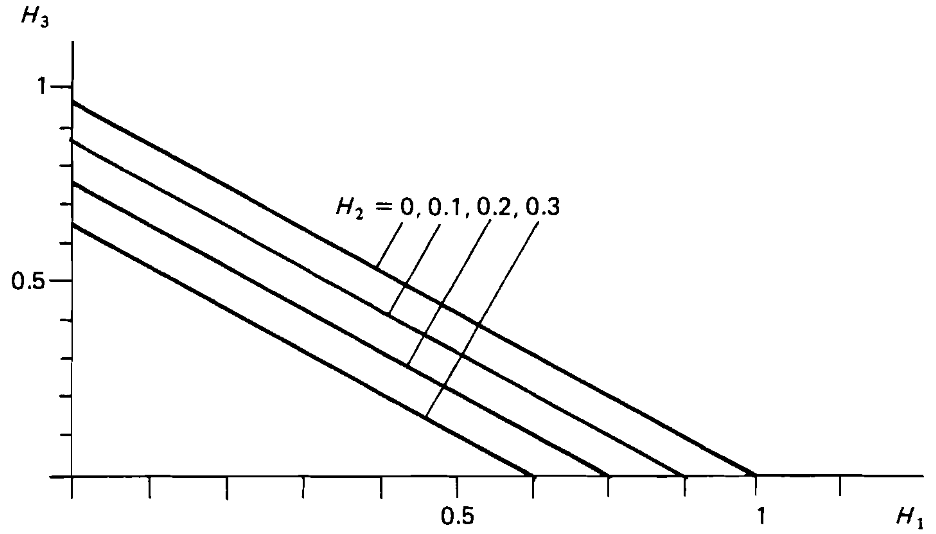


FIGURE 5 Boundaries of the stable niche (DC): "Big is beautiful" ($\sigma = +1$).

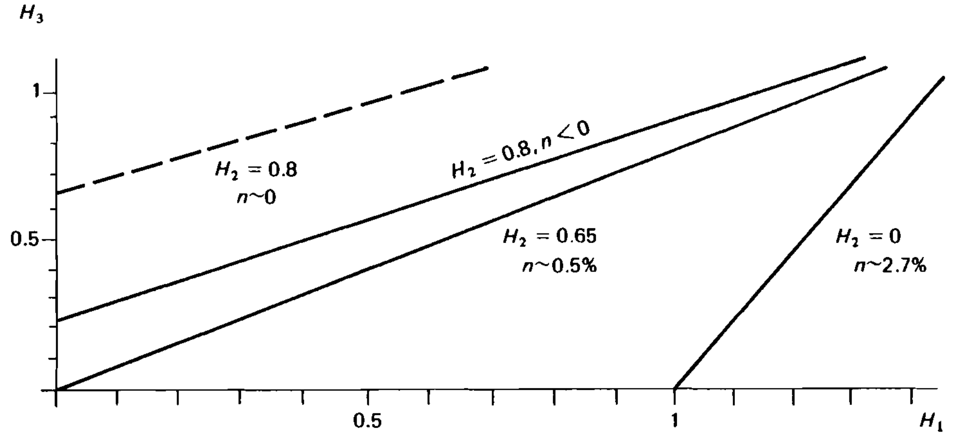


FIGURE 6 Boundaries of the stable niche (DC): "Small is beautiful" ($\sigma = -1$).

variable H_2 is perpendicular to the plane of the figure, and the intersections of the boundary surface with the planes $H_2 = 0, 0.1, 0.2$, and 0.3 (Fig. 5) and $H_2 = 0, 0.65$, and 0.8 (Fig. 6) are shown as solid lines. The region of stability always lies to the left of the respective lines. The growth rate along the line $H_2 = 0.8$ in Fig. 6 is negative; therefore, the actual boundary lies within the stable region. This is indicated by the dotted line, which represents the locus of zero growth.

Figure 7 shows a typical trajectory for our DC base case. The FP is located at $\epsilon_0 =$ 10 600 Wyr/smyr, $\kappa_0 = 16\,700$ \$yr/smyr, with a growth rate of $n_0 = 3.3$ percent. It will take five years to move along the trajectory from one square to the next.

7 TWO-COUNTRY INTERACTION

Thus far, the model has been used to describe an economy isolated from the rest of the world. Now, to be more realistic, we introduce two important links to other economies. First, most economies must import a significant part of their energy sources and must pay for it, with, for example, non-energy-related capital goods. Second, a rich economy may give away part of its ΔK – thereby reducing its growth rate, but not so much as to leave its favorable economic niche – to support a poor economy and thus induce its transition into a favorable niche.

We therefore modify the dynamics of eq. (16) in the following way:

– Reduce ΔK by a term proportional to E; the factor of proportionality is denoted by g. The numerical value of g is obtained heuristically:

  $g \doteq$ (fraction of energy imported) $\times$ (conversion factor bbl/Wyr) $\times$ (oil price \$/bbl) $\times$ $(1 -$ fraction recycled$)$ $= 0.3 \times 5 \times 10^{-3} \times 12 \times (1 - 0.7) \approx 5 \times 10^{-3}$ \$/Wyr.

  The last factor is included because petrodollars reinvested do not correspond to capital goods extracted from the system.

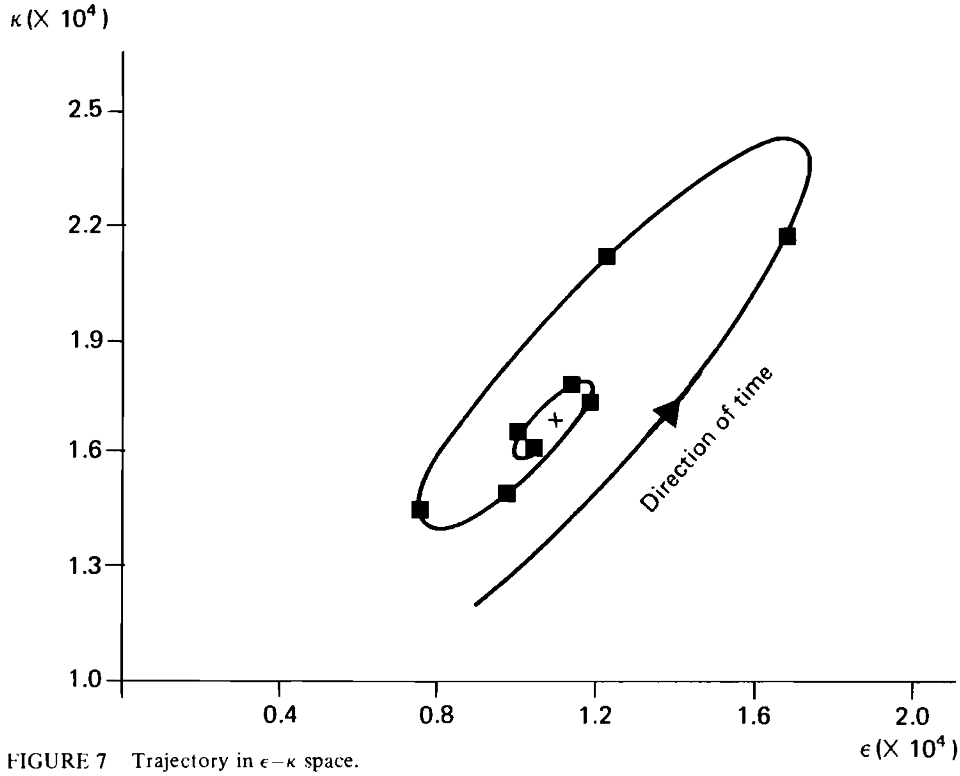


FIGURE 7 Trajectory in $\epsilon-\kappa$ space.

– Extract a fraction $r\Delta K$ from (or, for negative r, add a fraction $r\Delta K$ to) the total output of non-energy-related capital goods. This corresponds to foreign aid given away (or received).

In the spirit of the model, the phenomena of oil import and foreign aid are dealt with only schematically, by restricting discussion to the transfer of capital goods. r should not be confused with the well-known "0.7 percent of GNP" because foreign aid is measured on the scale of ΔK only.

In mathematical terms, we have to replace ΔK by $(1 - r)\Delta K - gE$; this is done simply by replacing eq. (14) by

$$\dot{X} = \mathbf{r}\mathbf{T}^{-1}\{(1-\alpha)X - X_{\mathrm{p}}\} - \mathbf{d}_g X \tag{37}$$

where

$$\mathbf{r} := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1-r & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathbf{d}_g := \begin{pmatrix} d_E & 0 & 0 \\ g & d_K & 0 \\ 0 & 0 & d_L \end{pmatrix} \tag{38}$$

For technical purposes, **r** may be incorporated into **T** by

$$\mathbf{T} \to \begin{pmatrix} a_{EE} & a_{EK}/1-r & a_{EL} \\ a_{KE} & a_{KK}/1-r & a_{KL} \\ a_{LE} & a_{LK}/1-r & a_{LL} \end{pmatrix}$$

Starting from the two base cases for a DC and an LDC economy, respectively, we may consider the resilience against variations of g and r (the base cases themselves correspond to $g = r = 0$). Given the price of energy (in terms of the oil price of 1976, 12 \$/bbl), how much foreign aid can the DC economy afford without leaving its favorable economic niche, decreasing its growth rate below zero, or both? How much foreign aid must be granted to an LDC economy so that both a transition from its unfavorable to a favorable economic niche and a positive growth rate are induced?

Figures 8 and 9 illustrate our findings. In Fig. 8, the solid line divides the stable region (on the left) from the unstable region (on the right); the broken lines are loci of constant

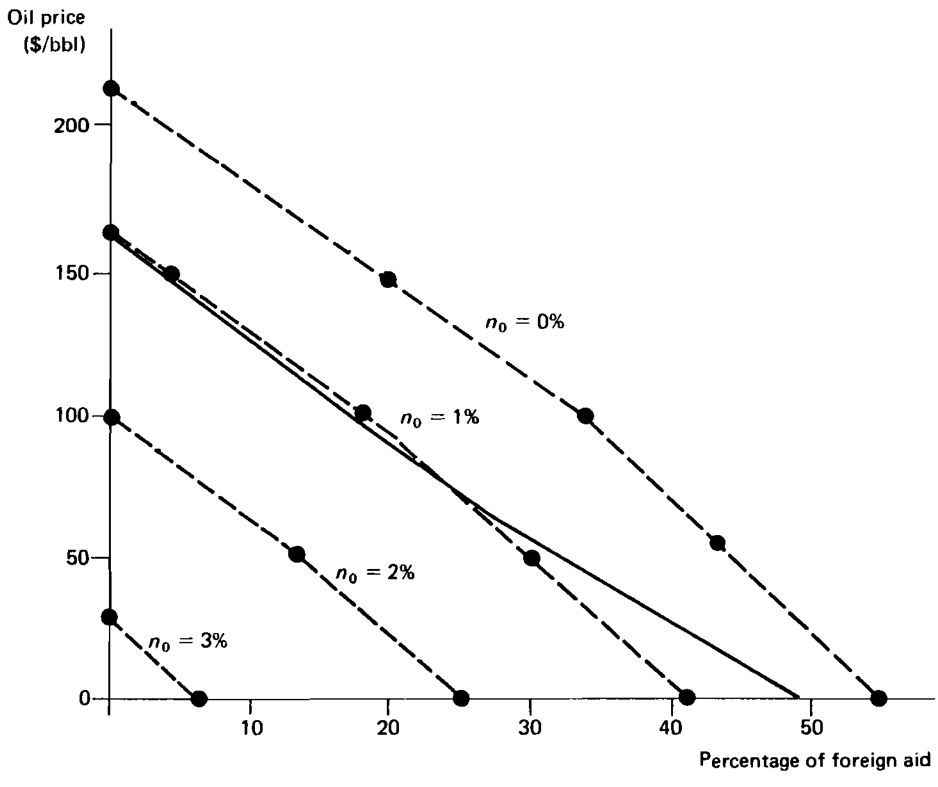


FIGURE 8 Economic aid: niche boundary (DC).

growth rate. In Fig. 9, the stable region is to the right of the boundary. In both cases, foreign aid is measured relative to ΔK of the respective country.

We may then combine the two economies by superimposing the two figures. Care must be taken to rescale foreign aid to one of the two economies. Hence, the ratio $\eta = \Delta K_{LDC}/\Delta K_{DC}$ must be given some value.* The result is displayed in Fig. 10(a) with

*The units on the abscissae in Figs. 8 and 9 are different; η, which relates capital goods production in the DC to that in the LDC, accounts for this difference.

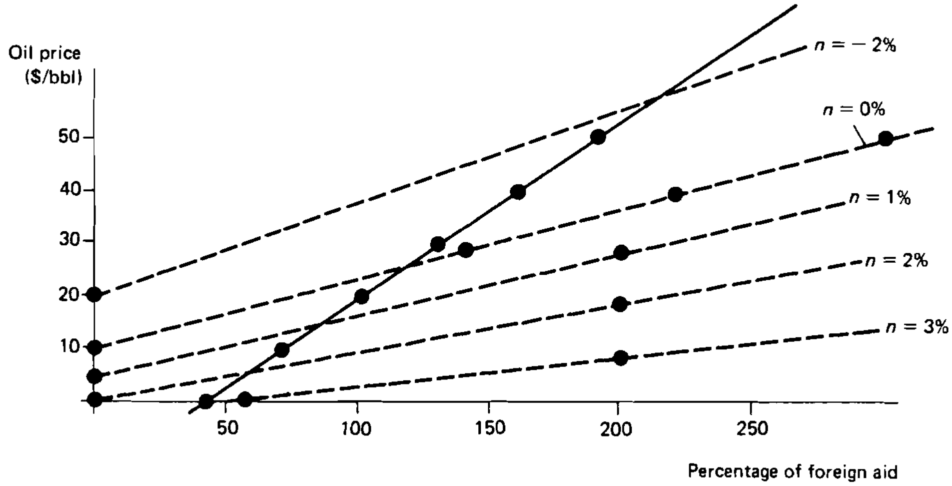


FIGURE 9 Economic aid: niche boundary (LDC).

η = 0.1 and in Fig. 10(b) with η = 0.2. Fig. 11 (of which Figs. 10a and 10b are actually slices) shows the situation with various values of η. As η is essentially a scale factor, diagrams with different values of η do not differ qualitatively.

8 OUTLOOK AND CONCLUSION

Two directions for further study come to mind immediately:

– Introducing an oil country as a full economy rather than merely as a sink for investment goods as in Section 7. Thus two-country (DC + oil country) or three-country (DC + LDC + oil country) interaction could be investigated. The existence of oil price thresholds (lower, upper, or both) for the stability of the oil country would be an interesting consideration.
– Relaxing our requirement of homogeneous production functions (neither economy nor diseconomy of scale). For this to be done meaningfully, E, K, and L must be reinterpreted as quantities referring to a "typical population" (say, 50 million inhabitants) or, better, to a "typical area" (say, 500 000 km^2; just as we introduced skill, we could weigh differently land of different productivity or other factors.* We could then multiply the linear production function for the investment sector with a "congestion function" or "agglomeration function," depending on the level of economic activity, which would be measured by a suitable linear combination of E, K, and L.

An interesting agglomeration function has the shape shown in Fig. 12. This congestion function expresses the unfavorable effects of levels that are either too high or too low. A system with this function would behave like a single-species ecological system with

*An FP in E, K, and L would imply that we predict the same equilibrium energy production for the US as for Liechtenstein, as long as they have the same parameters!

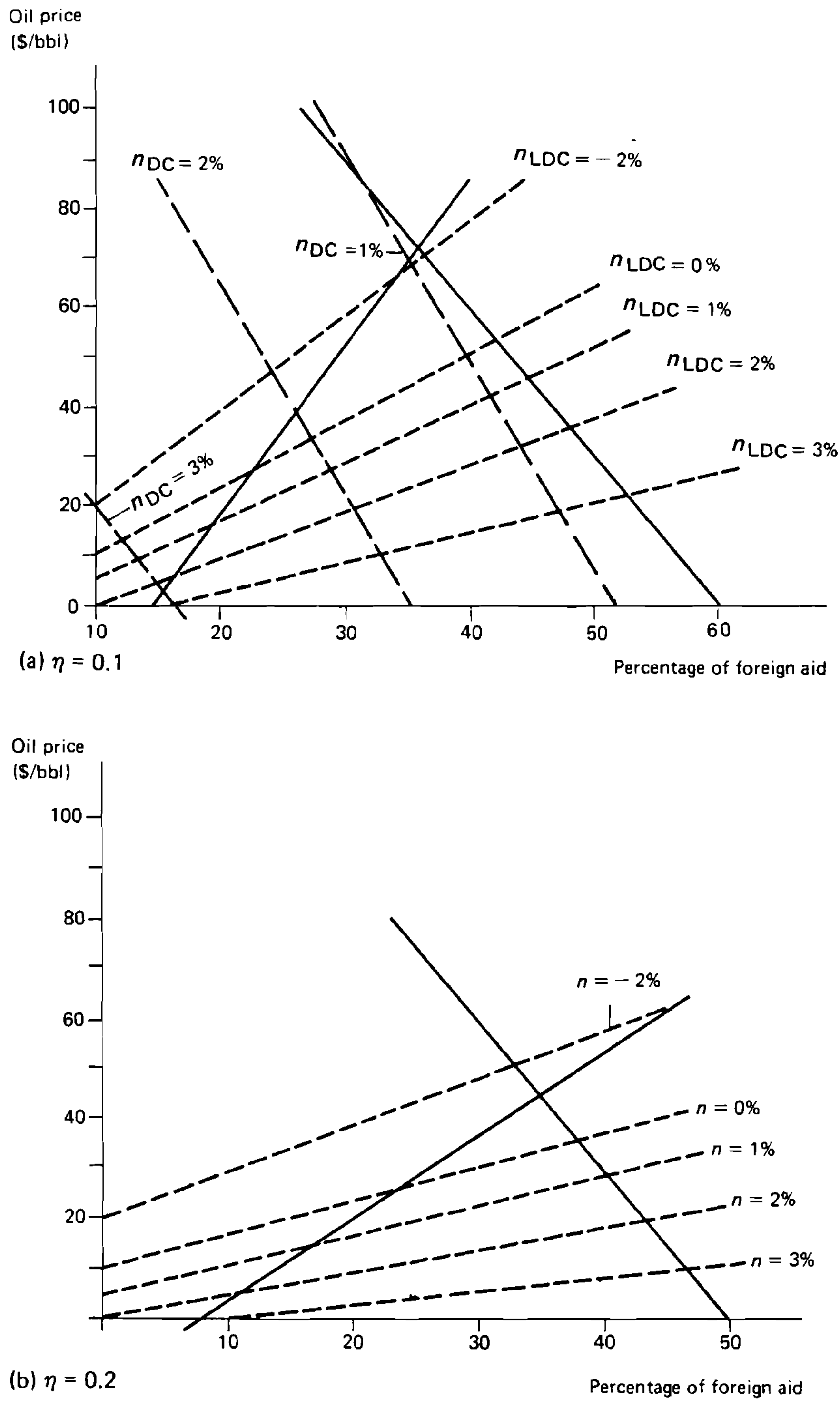


FIGURE 10 Maximum oil price.

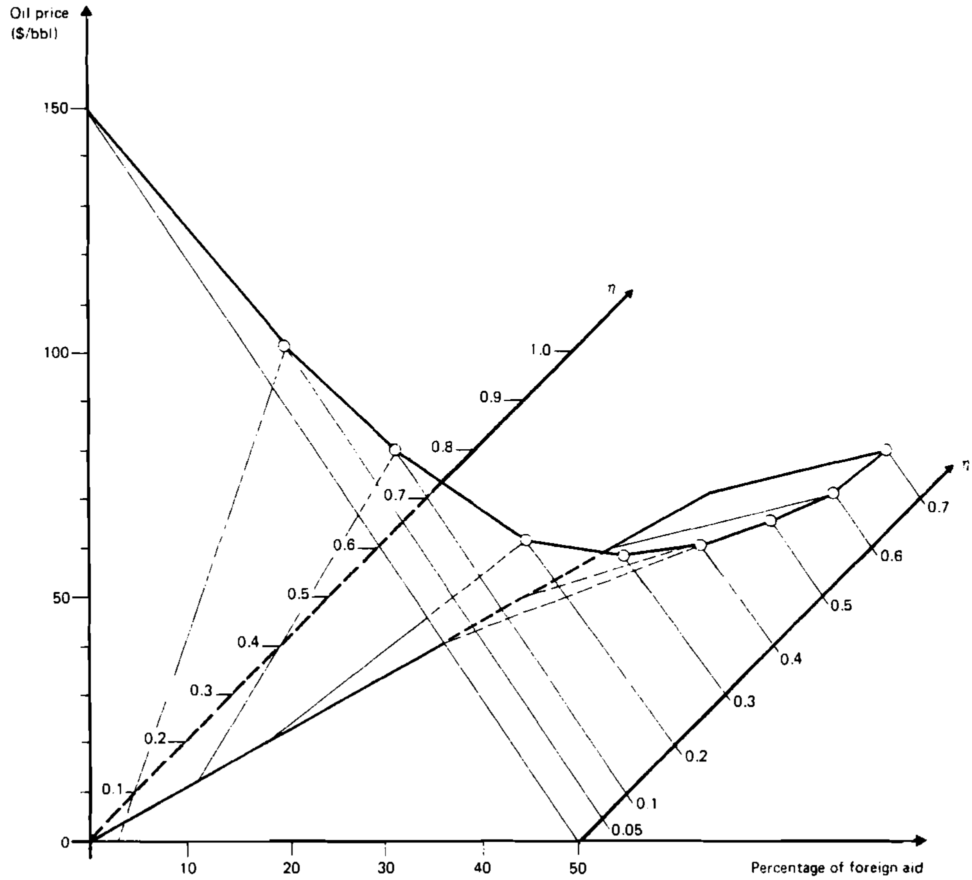


FIGURE 11 Maximum oil price variation with η.

the reproduction curve illustrated in Fig. 13 (Holling 1973), which shows a threshold level below which growth is negative and the system tends to zero and an equilibrium at a higher economic level; the trends are indicated by arrows. The behavior of the system transverse to the stable ray, i.e., in the ϵ–χ space, remains unaffected; the system thus has two FPs in E, K, and L outside of the origin, one of which can generate a separatrix.

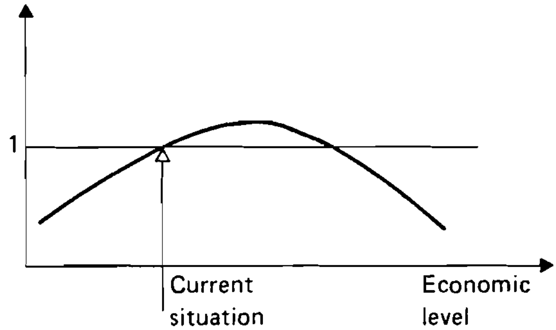


FIGURE 12 Congestion function.

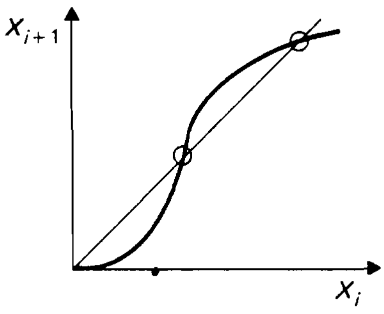


FIGURE 13 Reproduction curve.

The two-country interaction that we introduced must be considered as a first step. To help an LDC simply by pumping investment goods into its economy is certainly not sufficient: the transition to a favorable economic niche would be only a temporary one, and a cancellation (or perhaps merely a reduction) of foreign aid would result in an immediate reversal. The system parameters are only virtually, not intrinsically, changed by the exogenous support (as long as it is supplied). It is therefore necessary to introduce a coupling between the foreign aid and the parameters that corresponds to an actual improvement in the economy's infrastructure. As the economy improves, foreign aid may be reduced gradually until the transition from an LDC to a DC is completed. The mathematical way to express these mechanisms is unclear.

In summarizing our results, we must stress both what our approach can do and what it cannot do. Any detailed economic prediction – with quantitative results that inspire confidence – of course requires a much larger model. While it would be ridiculous to claim that an economy can be described accurately by just three state variables, we do feel justified in making three observations.

First, to the extent that the structure of our model (i.e., the choice of the state variables and the form and interrelation of the production functions) has something to do with an actual economy, we can deduce the existence of a slack-free region within the state space. As it can be said definitely that unpleasant situations will arise at the boundary (e.g., one or several outputs will tend to zero), proximity to the boundary of the slack-free region should be avoided.

Similarly, we have determined boundaries in parameter space (or, equivalently, in a space of scenario variables) across which the model's behavior changes drastically, showing instability, negative growth, and so forth. Because our knowledge is incomplete, parameter values close to those boundaries should be avoided as well (see Section 3).

Second, the model allows us to determine an FP (point of equal growth of E, K, and L) that shows one of the previously mentioned possible qualitative behaviors. These possible behaviors depend on the model economy's infrastructure, expressed in terms of a set of certain characteristic system parameters. The parameter space appears to be subdivided into cells, which we have called economic niches. By definition, all points within one niche correspond to economies with the same topological behavior. We have called niches with a stable associated FP favorable niches; their significance lies in the existence of a central region around the FP (within the slack-free region, of course) from which the slack-free region cannot be escaped. On the contrary, an economy starting outside the central region inevitably approaches the boundary of the slack-free region within a finite time. If such a trajectory were well off the boundary, a person "living" on it might not initially recognize a problem with the economy but, even with positive growth rates of E, K, and L, would observe that at least one of the outputs becomes zero. The initial conditions, i.e., the "right" amount of the available stocks, are essential for a favorable economic evolution.

Third, the two-country interactions studied in Section 7 model the structure to be expected for foreign aid given by an abstract DC to an abstract LDC, both of which are subject to the same oil price. An oil price sufficiently high inhibits any reasonable level of foreign aid; either the DC gives away too much or the LDC receives too little. It is a pleasant surprise that the limiting oil price comes out at the right order of magnitude – neither too close to the present level nor too high. A limiting price of, say, 10 000 $/bbl would make this feature irrelevant.

These qualitative results suggest interesting questions that we hope will be addressed through a full-size economic model.

ACKNOWLEDGMENTS

In preparing this paper, the authors benefited greatly from the collaboration of C. Riedel, from the continuous interest and guidance of W. Häfele, and from stimulating discussions with R. Bürk and H.-H. Rogner.

REFERENCES

Avenhaus, R., H.-R. Grümm, and W. Häfele (1975) New Societal Equations. Internal Paper. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Beaumol, W.J. (1977) Economic Theory and Operations Analysis. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Dorfman, R., P.A. Samuelson, and R.M. Solow (1958) Linear Programming and Economic Analysis. New York: McGraw-Hill Book Co.

Energy Systems Program Group of the International Institute for Applied Systems Analysis (1981) Energy in a Finite World: A Global Systems Analysis. W. Häfele, Program Leader. Cambridge, Massachusetts: Ballinger.

Grümm, H.-R., ed. (1975) Analysis and Computation of Equilibria and Regions of Stability with Applications in Chemistry, Climatology, Ecology, and Economics: Record of a Workshop. CP-75-8. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Grümm, H.-R. (1976) Definitions of Resilience. RR-76-5. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Grümm, H.-R., and L. Schrattenholzer (1976) Economy Phase Portraits. RM-76-61. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Häfele, W. (1975) Objective Functions. WP-75-25. Laxenburg, Austria: International Institute for Applied Systems Analysis. German version: Zielfunktionen, Beiträge zur Kerntechnik, Bericht der Gesellschaft für Kernforschung: 25–49. KFK 2200, Karlsruhe, FRG.

Häfele, W., and R. Bürk (1976) An Attempt of Long-Range Macroeconomic Modelling in View of Structural and Technological Change. RM-76-32. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Häfele, W., *et al.* (1976) Second Status Report of the IIASA Project on Energy Systems. RR-76-1. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Hicks, J. (1969) Capital and Growth. Oxford: Oxford University Press.

Holling, C.S. (1973) Resilience and Stability of Ecological Systems. RR-73-3. Laxenburg, Austria: International Institute for Applied Systems Analysis.
Holling, C.S., ed. (1978) Adaptive Environmental Assessment and Management. Chichester, UK: John Wiley and Sons.
Marchetti, C., and N. Nakicenovic (1979) The Dynamics of Energy Systems and the Logistic Substitution Model. RR-79-13. Laxenburg, Austria: International Institute for Applied Systems Analysis.
Samuelson, P.A. (1947, 1976) Foundations of Economic Analysis. Cambridge, Massachusetts: Harvard University Press and New York: Atheneum.

APPENDIX A Some Concepts of Dynamical Systems

We assume a deterministic system described by differential equations: knowing the state of the system at a particular time, one can calculate the time derivatives of all state variables. A "geometric" point of view is inherent in this approach: we introduce *state space*, each of whose points fully specifies a state of the system at one instant in time. The state space is spanned by the *state variables*. On the state space, we have a *time-evolution law*, possibly dependent on several parameters, ranging over *parameter space*. Under it, the states of the system move along *trajectories*. In our approach, we emphasize not so much a single trajectory as the structure of all trajectories. A *fixed point* (or *equilibrium*) is a state of the system that does not change in time; it may be stable or unstable.

Under general assumptions, the state space can be divided into *basins*, each consisting of states having a common future long-term behavior. Each basin contains one *attractor* representing this common behavior; it is the region in state space toward which all trajectories originating in the basin tend. The simplest attractor is a *stable fixed point* (or *stable equilibrium*); if the attractor of a basin is a stable fixed point and if the system starts in this basin, all state variables of the system will tend toward constant values. There are many more complicated types of attractors; the list is currently incomplete.

Basins are separated from each other by *basin boundaries* or *separatrices*. States on or very close to a basin boundary have uncertain futures because small modifications of the state variables may cause them to belong to different basins and thus to exhibit completely different long-term behavior.

The *phase portrait* is a full (at least qualitatively) description of the basins and attractors of a system. In general, it depends on the parameters of the system; qualitative changes of the phase portrait caused by parameters crossing certain boundaries are called *bifurcations*. These boundaries play a role similar to that of separatrices in state space.

The mathematical theory behind these concepts can be found in Arnold (1973), Grümm (1979), and Hirsch and Smale (1974).

APPENDIX B Dynamics Outside the Slack-Free Region

In looking at possible dynamics outside the slack-free region of the system defined in Section 4, we may rewrite eq. (14) as follows:

$$\Delta X = (\dot{X} + dX) = \mathbf{T}^{-1} X_{\mathrm{I}} \qquad X_{\mathrm{I}} = X(1-\alpha) - X_{\mathrm{p}} \tag{B1}$$

We have chosen this form to avoid the detailed structure of X_I, which is a known function of X. As discussed, if $\mathbf{T}^{-1}X_I \ngtr 0$, i.e., if some $(\mathbf{T}^{-1}X_I)_i < 0$, eq. (B1) does not make economic sense. We denote the value of ΔX obtained from eq. (B1) by ΔX_{vir} (the "virtual" gross production). To find a realistic ΔX, we turn again to eq. (6),

$$\Delta E = \min\left(\frac{E_{I,E}}{a_{EE}}, \frac{K_{I,E}}{a_{KE}}, \frac{L_{I,E}}{a_{LE}}\right) \tag{B2}$$

with similar equations for ΔK and ΔL. The assumption of "allocations without slacks," i.e., the equality of all terms in eq. (B2) and its ΔK and ΔL counterparts, leads to eq. (B1); thus this assumption can be fulfilled if and only if the system lies within the slack-free region. In this case, it leads to the unique dynamics contained in eq. (B1). Outside the slack-free region, we have to introduce slacks, taking into account that the terms in eq. (B2) and its counterparts will be equal. Summing the slacks occurring in the three sectors, we obtain

$$\Delta X = \mathbf{T}^{-1}X_W = \mathbf{T}^{-1}(X_I - X_S) \qquad 0 \leqslant X_S \leqslant X_I \tag{B3}$$

with X_S representing the stock of energy-related goods, non-energy-related capital goods, and skilled labor not allocated to production and X_W representing the stock actually "working" in the investment goods sectors.

As it stands, eq. (B3) of course does not define a unique evolution. We complement it by two requirements.

The first is the requirement of "Pareto optimality" of our allocation: no other allocation of E_N^I, K_N^I, and L_N^I to the three sectors leads to an increase in any of the quantities ΔE, ΔK, or ΔL without decreasing at least one of them. This is an obvious extension of the "allocation without slacks" possible within the slack-free region; there, this allocation is the unique Pareto-optimal one. Outside the slack-free region, there are generally several Pareto-optimal allocations.

The second is the requirement that uneconomical processes be shut off; if $\Delta E_{vir} < 0$, the realistic ΔE is set to zero. The allocation without slacks that, if possible, would give the virtual ΔE would require the ΔE production to run backwards; as this is not possible, we handle the situation by shutting off ΔE production completely. This requirement follows from the first if we have only two production functions; one can argue for it using familiar arguments from linear programming.

Both requirements might still fail to define a unique evolution of our system by giving unique expressions for $\dot{E}$, $\dot{K}$, and $\dot{L}$. We call any evolution of the system outside the slack-free region fulfilling eq. (B3) and these two requirements a rational evolution.

To illustrate the general situation, we assume that the system is just crossing from inside the boundary of the slack-free region corresponding to $\Delta E = \dot{E} + d_E E = 0$. We write for the "real" evolution outside the slack-free region

$$\begin{aligned} &\dot{E} + d_E E = 0 \\ &a_{KE}(\dot{K} + d_K K) + a_{LE}(\dot{L} + d_L L) = E_I - E_S \\ &a_{KK}(\dot{K} + d_K K) + a_{LK}(\dot{L} + d_L L) = K_I - K_S \\ &a_{KL}(\dot{K} + d_K K) + a_{LL}(\dot{L} + d_L L) = L_I - L_S \end{aligned} \tag{B4}$$

together with

$$0 \leqslant E_S \leqslant E_I \qquad 0 \leqslant K_S \leqslant K_I \qquad 0 \leqslant L_S \leqslant L_I \tag{B5}$$

and the Pareto optimality of the allocation. At least one of the slacks must obviously be zero; otherwise, we could increase ΔK, say, without decreasing ΔL.* If one slack is zero, the other two are linearly related; taking the inequalities of eq. (B5) into account, we see that for all rational allocations $X_S = (E_S, K_S, L_S)$ must lie on one of three straight segments. As eq. (B4) is an affine relation between $(\dot{K}, \dot{L})$ and X_S, all rational allocations lead to the following net increases:

$$\dot{E} = -d_E E$$
$$\begin{pmatrix} \dot{K} \\ \dot{L} \end{pmatrix} \in S \tag{B6}$$

S denotes the set in $(\dot{K}, \dot{L})$ space that is the image of the previously mentioned straight segments. If S contains a "greatest point," i.e., if both $\dot{K}$ and $\dot{L}$ are larger than at any other point of S, there is only one rational allocation, and the time derivatives $\dot{E}$, $\dot{K}$, and $\dot{L}$ are uniquely determined; if not, the system will have a residual freedom. The two situations in $(\dot{K}, \dot{L})$ space (S is the boundary of the triangle) are shown together with the virtual time derivatives in Fig. B1. Fig. B1(a) shows a unique evolution; in Fig. B1(b), any point on the upper right-hand segment corresponds to a rational evolution; nevertheless, all rational trajectories will lie inside a certain "fan."

This ambiguity is not serious: as we come close to the slack-free region, S becomes small and contracts to $(\dot{K}, \dot{L})_{vir}$ as we reach the boundary. If the virtual evolution is stable, as analyzed in Section 5, its trajectories leaving the slack-free region will come back. As indicated in Fig. B1, $\dot{E}_{vir} < \dot{E}, \dot{K}_{vir} > \dot{K}, \dot{L}_{vir} > \dot{L}$ for all rational allocations (if the trajectory crosses the $\Delta E = 0$ boundary). Outside the slack-free region a "rational"

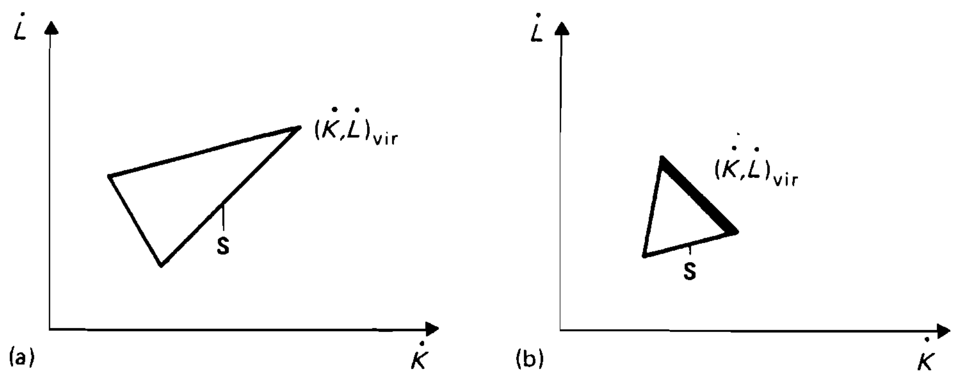


FIGURE B1 Rational evolutions.

*Equation (B4) of course holds on the boundary of the slack-free region, too, with $E_S = K_S = L_S = 0$. Furthermore, the "real" $\dot{K}$ and $\dot{L}$ are the same as the virtual $\dot{K}_{vir}$ and $\dot{L}_{vir}$ defined by eq. (B1). Thus the evolutions inside and outside the slack-free region fit together continuously.

trajectory will of course not coincide with the "virtual" trajectory but, owing to these inequalities, it will lie closer to the slack-free region than will the "virtual" trajectory. Thus any rational evolution leads the system back to the slack-free region, just as the "virtual" trajectory does.

APPENDIX C A Criterion for the Number of Fixed Points

The criterion explained here distinguishes between the two alternatives discussed in Section 5. We write eq. (18′) in the form

$$\det\,[\mathbf{1} - \alpha - (d + n)\mathbf{T}] = 0 \qquad \text{(C1)}$$

which yields ($x = d + n$)

$$A + BX + CX^2 + DX^3 = 0 \qquad \text{(C2)}$$

Here $A = \det(\mathbf{1} - \alpha)$, $D = -\det \mathbf{T}$, and B and C are defined correspondingly. Substituting

$$p := B/D - C^2/3D^2 \qquad q := 2C^3/27D^3 - BC/3D^2 + A/D \qquad y := X + C/3D \qquad \text{(C3)}$$

into the normal form of an equation of third degree,

$$y^3 + py + q = 0 \qquad \text{(C4)}$$

brings us back to eq. (C3). Theory now tells us that there will be one real solution of eq. (22) if

$$q^2/4 + p^3/27 > 0 \qquad \text{(C5)}$$

and three real solutions if

$$q^2/4 + p^3/27 < 0 \qquad \text{(C6)}$$

Substitutions of eq. (C3) into eq. (C5) or eq. (C6) produce a hypersurface in parameter space that separates the two alternatives. Unfortunately, this relation is too clumsy to be written down explicitly.

Equations (C5) and (C6) are only a criterion for the simplified case (equal ds and Xp = 0); the hypersurface mentioned will be continuously deformed during the transition to the general case.

REFERENCES TO APPENDIXES

Arnold, V.I. (1973) Ordinary Differential Equations. Cambridge, Massachusetts: MIT Press.

Grümm, H.-R. (1979) Introduction to Dynamical Systems. Internal Paper. Laxenburg, Austria: International Institute for Applied Systems Analysis.

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

INNOVATION AND EFFICIENCY

Heinz-Dieter Haustein, Harry Maier, and Luitpold Uhlmann
International Institute for Applied Systems Analysis, Austria

SUMMARY

Innovation is a complex phenomenon that involves all spheres of technological, economic, and social activity, from research and development to investment, production, and application. In the management of innovation the relationship between innovation and efficiency is the key issue. In this report, therefore, we elaborate on a method for measuring efficiency in the innovation process. The core of our concept of efficiency is the link between the efficiency of the production unit that has adopted an innovation (dynamic efficiency) and the efficiency of the entire production field within which production units must act (average efficiency). The development of relative efficiency is connected to differences between basic, improvement-related, and pseudo innovations and to the decision-making environment for managers.

Factors influencing innovative activities follow a continuum of efficacy ranging from inhibiting to strongly promoting innovative activities. Looking at the innovation process from the standpoint of the innovating system, we distinguish major determinants of performance and then compare the performance of industrial organizations through a profile showing these determinants in research and development, production, and marketing and in management at all stages.

---

1 MEASURING EFFICIENCY IN THE INNOVATION PROCESS

1.1 Principal Indicators of Efficiency

Before presenting our model of the innovation process, we would like to describe the economic environment of innovations; without knowing the needs of and possibilities offered by this environment, one cannot understand the mechanism of technological change. The results of interactions between innovations and their environment are usually measured in terms of economic efficiency. In this report, therefore, we focus our attention on the problem of efficiency.

The measurement of efficiency in socioeconomic and technical–economic processes is a wide and comprehensively explored field. We differentiate in this report among technical efficacy, economic efficiency, and social effectiveness. Specific measures of technical efficacy are clearly defined and verifiable, but it is difficult to give general indicators for the technical efficacy of such products as automobiles, washing machines, and television sets. This generalization is even more true for measures of economic efficiency, which are by definition more aggregate than are technical indicators. Here we also encounter other problems: the difficulty of clearly adjoining elements to defined sets, the complicated procedure of statistical inquiry, and the lost contact between user and producer of data. Yet the measurement of social effectiveness is the most complicated, as social welfare and social climate cannot be measured successfully by the monetary indicators that are so useful in economics.

Innovation is a complex phenomenon that involves all spheres of technological, economic, and social activity, from research and development to investment, production, and application. In the early stages there are only two general indicators of innovative efficiency, which can be evaluated and predicted in rough variants (see Fig. 1). These are the level of technology and the desired range of application. These indicators are combined into certain coefficients and are connected with recognized needs, time limitations and competitive pressures, and available resources. The level of technology and range of application determine the compatibility or interference with existing equipment and skills, degree of interdependence, degree of complexity, and scale. For these coefficients we need additional information that is not available during the first stages of research and development. As the innovation process progresses, however, we are able to calculate the risk factor, development time, lifetime, and resource requirements. We should then gradually make the previously mentioned coefficients more precise. Later, we can calculate in monetary measures the economic benefits and expenditures and can determine other indicators of economic and social efficiency.

Owing to the interference of the new technology with existing equipment and skills, however, it is not easy to isolate the efficiency of the innovation from that of the production unit introducing the new technology. The only available solution to this problem is to compare an innovating unit with a noninnovating unit, but neither the results of interference with existing equipment and skills nor the effects of new elements can be isolated.

It is difficult enough to measure efficiency in comparing similar industries or countries, but we encounter many more problems in trying to compare those under different social systems; both the goals and underlying mechanisms of socioeconomic actions and the reference system for measuring efficiency are different. Table 1 suggests that, at least for some indicators, there are no great differences between market and planned economies. We must ensure, however, that similar indicators are used for different goals in both systems and that in planned economies these indicators are calculated in a uniform way within the planning process connecting all levels from the plant to the national economy. A common reference system is needed and is plausible primarily

- In fields involving such cooperative action as trade, exchange of technologies, and investigation of solutions to world problems
- At the level of intermediate goals

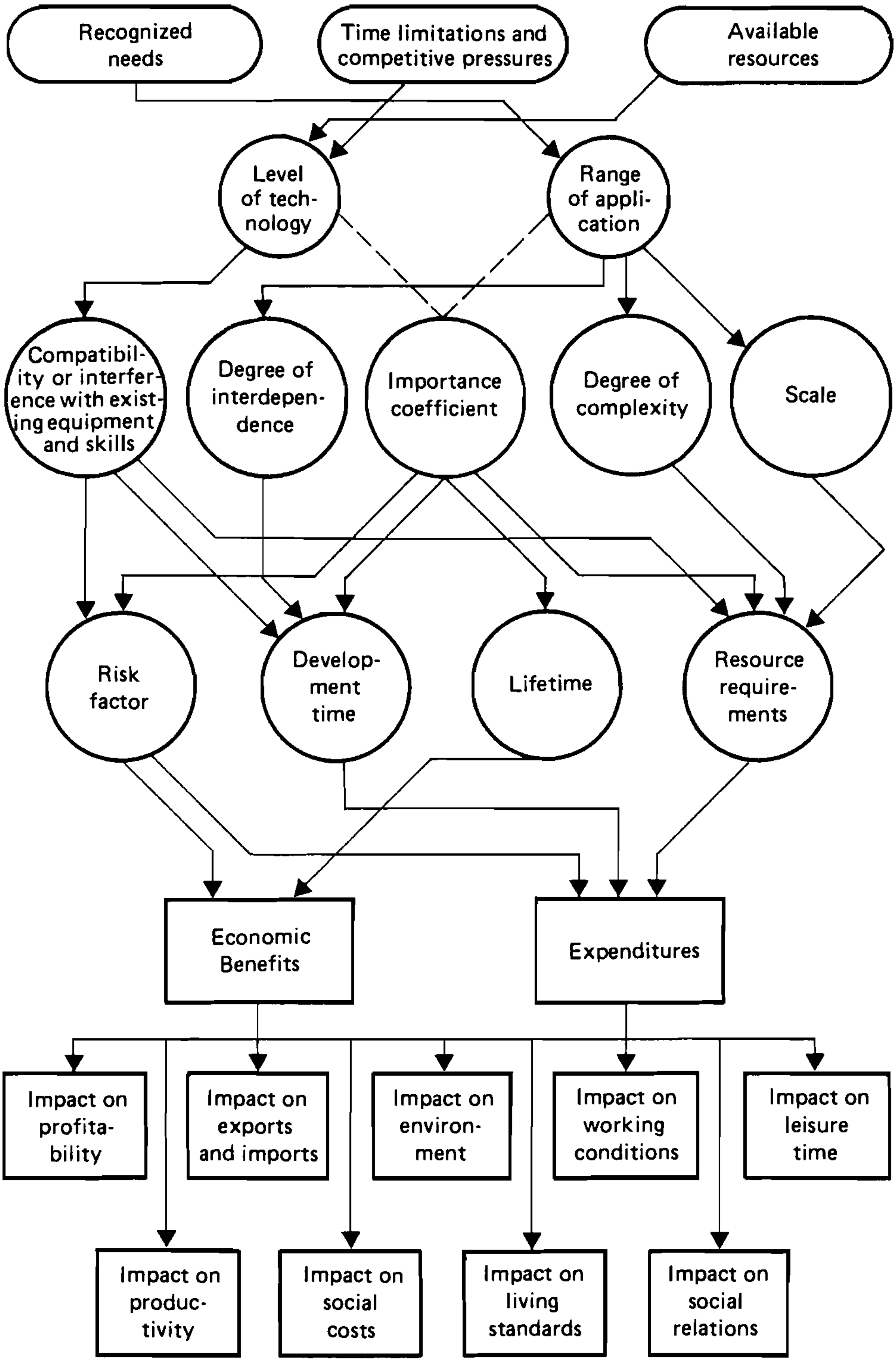


FIGURE 1 Indicators of innovative efficiency.

TABLE 1 Measures of efficiency in market and planned economies at the company and national levels.

	Measures of efficiency	
Level	Market economy	Planned economy
Company	Growth rate (sales and profits)	Growth rate (net product)
	Productivity (labor and capital)	Productivity (labor)
	Return on book value	Return on funds
	Profit margin (as percent of sales)	–
	Earnings per share	–
	Market share	–
	–	Export profitability
	–	Cost factor
	–	Material intensity of production
	–	Capital coefficient (output per unit of funds)
National	Growth rate (national income)	Growth rate (national income)
	Productivity (labor)	Productivity (labor)
	Balance of payments	Balance of payments
	Capital coefficient	Capital coefficient

One of the most important intermediate goals in both kinds of economy is productivity. It is generally accepted that productivity growth rates over a long period reflect the true economic performance of an industry or of a nation. Data on productivity growth rates are available in all countries and are more comparable than are indicators of profitability. The development of labor productivity could be an important indicator of a country's technological innovativeness, but we must also take into account the constraints connected with this indicator.

$$\text{Labor productivity} = \frac{\text{Gross product}}{\text{Number of employees}} \text{ or } \frac{\text{Net product}}{\text{Working hours}}$$

Statistical details show that the gross domestic product (GDP) is not the same in the Organisation for Economic Co-operation and Development (OECD) and Council for Mutual Economic Assistance (CMEA) countries. CMEA countries include material input from outside the firm, while OECD countries do not. On the other hand, the figures of CMEA countries include only goods and so-called productive services – not banking or insurance operations, rent, and similar factors. Figure 2 shows the principal similarities and differences in methodology, while Table 2 gives a practical example. The net product according to the methodology of planned economies is 20 to 30 percent lower than the same net product according to the methodology of market economies. On the level of the industry, the methodologies are more similar, and the production value includes sales and the changes in inventories of intermediate products. We also find differences in methodology with respect to the number of employees; while apprentices are included as employees in OECD countries, they are not in CMEA countries.

We cannot, therefore, expect the official productivity statistics of OECD and CMEA countries to give us a complete picture. However, the differences counteract and neutralize each other in part; this is particularly evident if we investigate growth rates. In Table 3 we present industrial productivity growth rates in major developed countries for 1963–1973, 1973–1977, 1978, and 1979. Figure 3 shows the decline in productivity growth rates for

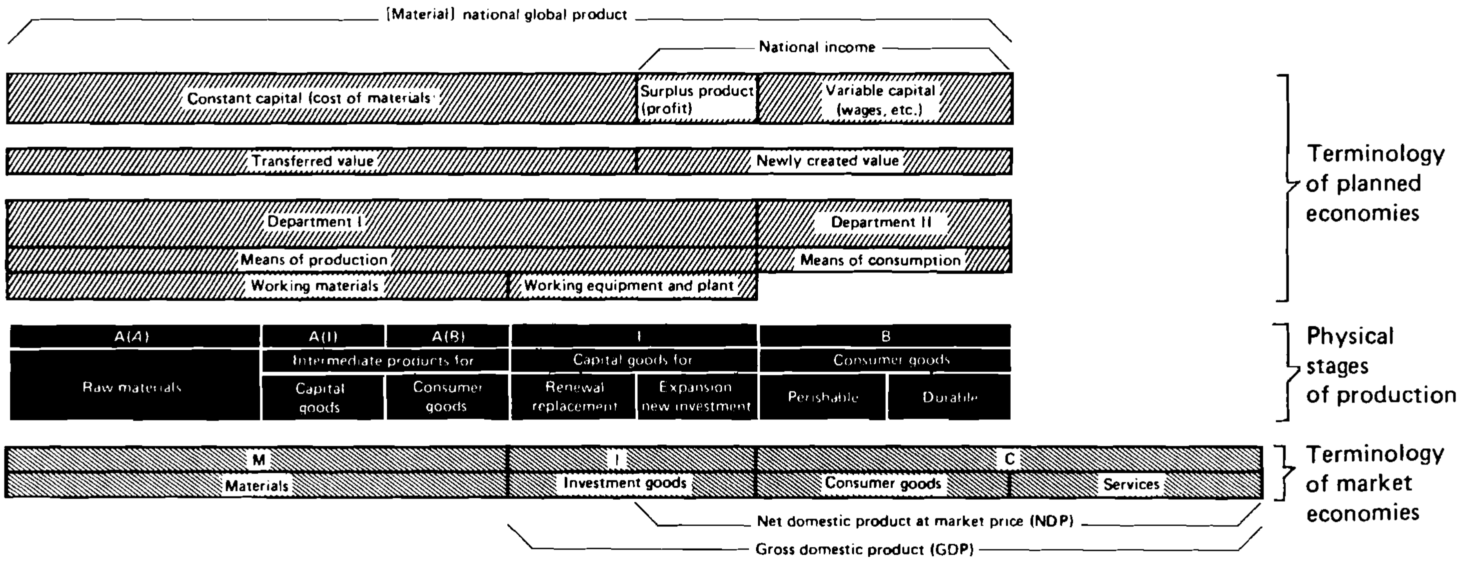


FIGURE 2 Content and mutual relations of Eastern and Western macroeconomic concepts in the sphere of production. Source: Eisendraht 1964.

TABLE 2 Comparison of national income per inhabitant and national income in the USA, USSR, FRG, and Japan, 1977.

	National income per inhabitant						National income			
	According to the methodology of market economies, including nonproductive sector (services)		According to the methodology of planned economies, excluding nonproductive sector (services)							
	At official exchange rate		At official exchange rate		At purchasing power value		At official exchange rate		At purchasing power value	
Country	Dollars	Percent	Dollars	Percent	Dollars	Percent	Billion dollars	Percent	Billion dollars	Percent
USA	7010	100	4655	100	4655	100	1010	100	1010	100
USSR	–	–	2115	45	2599	56	548	54	673	67
FRG	4480	64	3270	70	2265	48	196.1	19	135.8	13
Japan	3020	43	2235	48	–	–	242	24	–	–

SOURCE: Statistical Yearbook of the USSR 1977.

TABLE 3 Industrial productivity growth rates in major developed countries, 1963–1979.

Country	Industrial productivity growth rate 1963–1973	Industrial productivity growth rate 1973–1977	Change in industrial productivity growth rate	Change in output growth rate	Industrial productivity growth rate 1978	Industrial productivity growth rate 1979
Planned economies						
USSR	5.6	4.8	–0.8	–1.4	3.5	2.4
Poland	5.9	8.0	2.1	3.6	4.8	3.3
GDR	5.3	5.3	0	–0.3	4.2	4.2
Czechoslovakia	5.4	5.6	0.2	–0.7	4.1	2.9
Hungary	4.6	6.3	1.7	–0.2	5.9	4.5
Bulgaria	6.7	6.7	0	–4.3	6.4	–
Rumania	7.0	7.8	0.8	0.1	6.8	6.4
Market economies						
USA	2.1	1.0	–1.1	–3.5	1.9	1.3
Japan	8.9	3.7	–5.2	–9.5	8.8	9.6
FRG	5.3	3.6	–1.7	–4.4	2.3	4.0
France	5.2	4.0	–1.2	–3.4	5.0	–
UK	3.9	1.3	–2.6	–3.6	3.4	3.1
Canada	3.6	0.8	–2.8	–4.4	4.7	2.5
Italy	5.6	0.8	–4.8	–4.1	3.0	–

the economy of the FRG for the 27 years from 1951 through 1977. The average annual decline in productivity growth for this period was 0.2 percent.

According to a recent study (OECD Economic Outlook 1979), OECD countries are alarmed about their continuing decline in industrial productivity growth rates in the seventies. The productivity growth rates of the seven major CMEA countries are higher

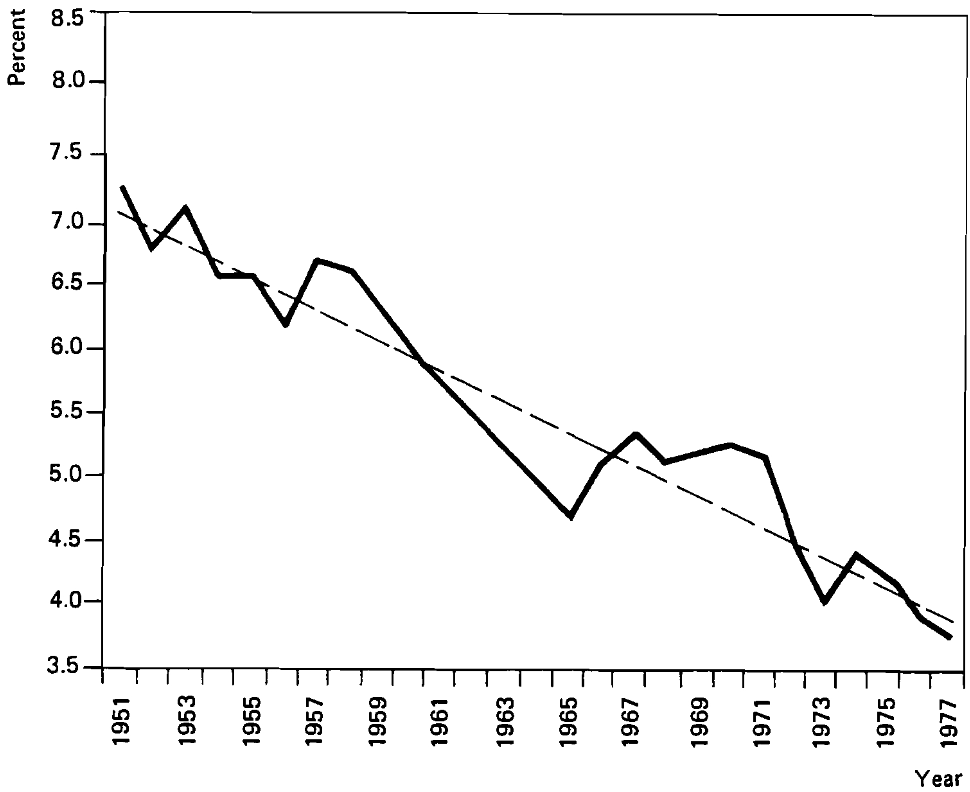


FIGURE 3 Productivity growth rates of the national economy of the FRG, 1951–1977 (moving averages of five years for gross domestic product in 1970 prices per working hour, all employees). Dotted line is trend line.

than are those of the seven major OECD countries. We do, however, find a negative or zero change in the productivity growth rates of the USSR, Czechoslovakia, and the GDR for the 1973–1977 period. Looking at data from several industries in Table 4, we note a decline in the productivity growth rates of nearly all industries in several countries. Poland, which had rapid industrialization during the reference period, is the single exception. Yet industrial productivity growth, which remains the main source of national welfare and the prime contributor to international competitiveness and equalization of gaps in resources, is important in both less developed and developed countries. Planned economies are seeking to reduce the productivity gap in order to be at the same level as market economies. The time needed for equalization of productivity levels depends on the size of the gap, current growth rates, and future change in growth rates. Appendix A presents a method for calculating the time needed to equalize productivity levels in two countries.

The present decline in productivity growth rates, which is of course not conducive to equalizing productivity levels, cannot be explained simply by the levels of productivity reached. Instead, there must be a cause having a similar effect in all countries. The lack of basic innovations might be such a universal factor. The most important growth industries of the last 30 years have been chemicals, electrical engineering, automobiles, plastics, petroleum products, and aircraft. Now, however, we see a negative change in productivity growth even in these industries – which recently have not been compensated by new basic

TABLE 4 Productivity growth rates of several industries in various countries[a] (market and planned economies), 1963–1977.

Industry[b]	USA			Japan			FRG			UK		
	1963–1973	1973–1977	Change	1963–1973	1973–1977	Change	1963–1973	1973–1977	Change	1963–1973	1973–1977	Change
1a. Food and tobacco 1b. Food	2.5	3.3	0.8	5.8	−0.4	−6.2	5.3	6.1	0.8	3.2	1.6	−1.6
2a. Textiles 2b. Textiles without clothing	3.4	2.7	−0.7	8.0	8.3	0.3	6.3	7.2	0.9	6.0	−0.1	−6.1
3a. Pulp, paper, and paper products 3b. Paper-making	4.6	−1.0	−5.6	9.8	2.9	−6.9	6.1	5.3	−0.8	3.9	−0.8	−4.7
4a. Chemicals 4b. Chemicals rubber, and asbestos	4.6	0.6	−4.0	11.3	5.4	−5.9	9.0	3.0	−6.0	6.9	2.1	−4.8
5a. Petroleum and coal products 5b. Fuel and production of fuel from coal, oil, and shale	3.7	0.4	−3.3	9.5	−0.9	−10.4	4.0	3.0	−1.0	6.9	−2.2	−9.1
6a. Nonmetallic mineral products 6b. Construction materials	1.5	0.5	−1.0	7.0	−2.2	−9.2	6.0	7.1	1.1	5.3	−0.8	−6.1
7a. Basic metal 7b. Ferrous metals (including ore extraction)	1.7	−3.6	−5.3	13.2	−1.0	−14.2	6.4	−0.5	−6.9	2.9	−3.3	−6.2
8a. Processed metal products 8b. –	1.9	0.2	−1.7	10.4	−1.4	−11.8	4.7	5.2	0.5	1.8	0.1	−1.7
9a. Machinery 9b. Engineering and metalworking	2.2	−0.7	−2.9	9.0	6.3	−2.7	3.9	2.8	−1.1	4.2	−1.4	−5.6
10a. Electrical machinery, equipment, and supplies 10b. –	4.8	1.4	−3.4	12.5	11.1	−1.4	6.5	7.3	0.8	5.3	0.3	−5.0
11a. Transport equipment 11b. –	2.9	3.6	0.7	9.2	10.8	1.6	3.5	2.7	−0.8	2.3	−2.4	−4.7
12a. Precision instruments 12b. –	2.6	1.3	−1.3	6.0	14.9	8.9	4.8	3.7	−1.1	6.5	−2.8	−3.7
13a. – 13b. Timber and woodworking												
14a. – 14b. Glass, china, and pottery												
15a. – 15b. Printing												

Industry[b]	USSR			Czechoslovakia			GDR			Poland		
	1963–1973	1973–1977	Change	1963–1973	1973–1977	Change	1963–1973	1973–1977	Change	1963–1973	1973–1977	Change
1a. Food and tobacco												
1b. Food	4.0	3.0	−1.0	3.6	3.5	−0.1	4.0	2.3	−1.7	2.5	4.2	1.7
2a. Textiles												
2b. Textiles without clothing	4.9	3.1	−1.8	4.7	4.9	0.2	6.8	6.3	−0.5	4.5	8.7	4.2
3a. Pulp, paper, and paper products												
3b. Paper-making	5.8	4.5	−1.3	4.0	5.9	1.9	5.6	5.0	−0.6	3.4	8.9	5.5
4a. Chemicals												
4b. Chemicals, rubber, and asbestos	7.3	6.8	−0.5	7.3	7.2	−0.1	6.5	7.6	1.1	8.5	9.8	1.3
5a. Petroleum and coal products												
5b. Fuel and production of fuel from coal, oil, and shale	6.9	4.5	−2.4	6.3	2.5	−3.8	6.2	3.6	−2.6	5.1	4.9	−0.2
6a. Nonmetallic mineral products												
6b. Construction materials	6.2	4.0	−2.2	5.2	5.8	0.6	6.0	4.7	−1.3	5.9	7.3	1.4
7a. Basic metal												
7b. Ferrous metals (including ore extraction)	–	–	–	5.2	4.1	−1.1	7.1	5.3	−1.8	6.2	9.5	3.3
8a. Processed metal products												
8b. –												
9a. Machinery												
9b. Engineering and metalworking	7.9	7.7	−0.2	6.8	7.0	0.2	6.0	5.6	−0.4	7.9	10.8	2.9
10a. Electrical machinery, equipment, and supplies												
10b. –												
11a. Transport equipment												
11b. –												
12a. Precision instruments												
12b. –												
13a. –												
13b. Timber and woodworking	4.9	3.9	−1.0	4.9	6.7	1.8	6.1	5.5	−0.6	4.0	11.2	7.2
14a. –												
14b. Glass, china, and pottery	7.9	6.6	−1.3	4.9	6.7	1.8	5.3	5.9	0.6	5.9	12.0	6.1
15a. –												
15b. Printing	–	–	–	5.9	5.6	−0.3	5.6	3.1	−2.5	4.2	12.9	8.7

[a] Abbreviations used here and elsewhere in this report are as follows: United States of America (USA), Federal Republic of Germany (FRG), United Kingdom (UK), Union of Soviet Socialist Republics (USSR), and German Democratic Republic (GDR).

[b] As industries in market and planned economies are not strictly comparable, we have indicated differences by dividing each and lettering the resulting divisions (a or b); in some cases (indicated by –), no counterpart exists.

SOURCE: Adapted from OECD Economic Outlook 1979 and statistical yearbooks of CMEA countries.

innovations. How, then, might the lack of basic innovations explain the decline in productivity growth rates?

Two tendencies have a great effect on efficiency. First, an increasing capital coefficient leads toward improvement of a given technological system. Essential changes are of no interest if they are linked with large losses in capital funds, and the capital coefficient is a general measure for many specific problems at the level of the firm. Table 5 shows some of the problems arising at this level (in marketing, production, research and development, management, and social consequences) during the transition from a policy of improvement (that is, changes of lower order) to one of basic technological change. Second, many

TABLE 5 Implications of policy of improvement or of basic technological change at the level of the firm.

	Implications of policy	
Factor	Improvement	Basic technological change
Marketing	Demand relatively low, well known, and predictable	Demand high and relatively unpredictable
	Risk of failure low	Risk of failure high
	Acceptance rapid	Acceptance slow initially
	Well-known marketing channels used	Creation of a new marketing system necessary
Production	Capacities of existing labor, skills, and cooperation used maximally	Capacities of existing labor, skills, and cooperation becoming obsolete
	Learning processes and designs streamlined	Learning processes interrupted
	Risk in quality and process planning high	Problems in quality, costs, and effects new and unanticipated
Research and development	Existing research and development potential used	Advanced research potential needed
	Basic research not needed	New research fields and disciplines needed
	Research and development risk relatively predictable	Research and development risk high
Management	Familiar management systems used and given organizational solutions adapted	New management skills, methods, and organizational solutions needed
		Complexity increased
Social consequences	Unpredictable problems rare or nonexistent	Legal and social acceptance unpredictable

firms show a strong tendency to follow a policy of improvement. Figure 4 and Table 6 show this development over a 20-year period in the USA, where the number and percentage of radical breakthroughs are declining rapidly. The same situation can be identified in other countries.

On the other hand, the situation changes according to the industry or group of products. Over the 1953–1973 period, the number of major innovations in electrical equipment

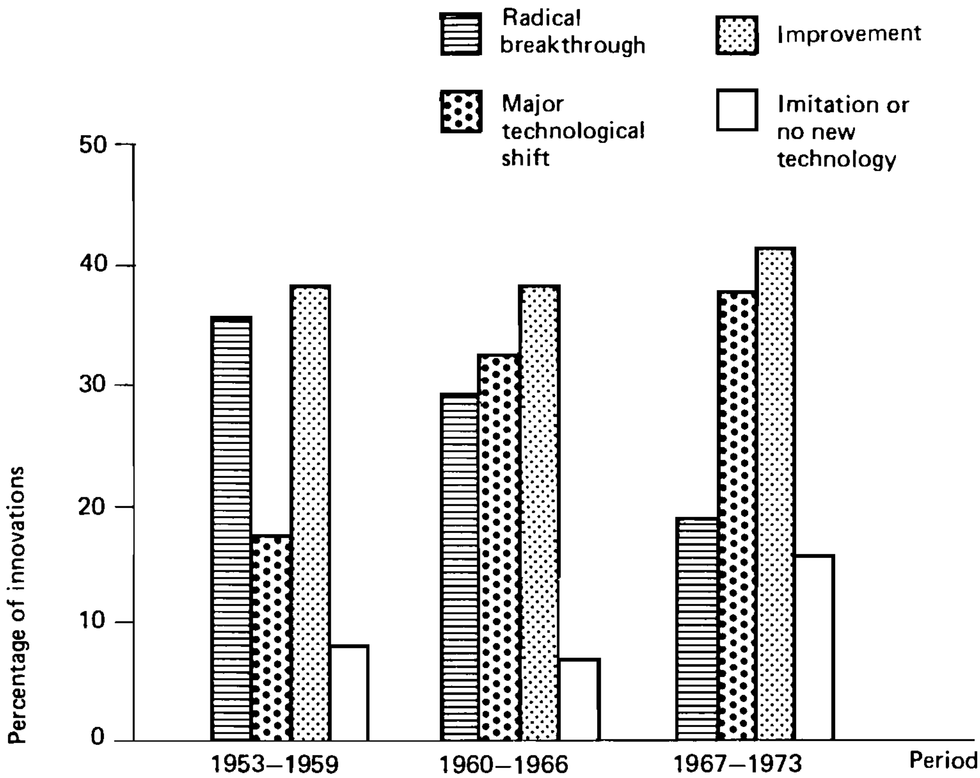


FIGURE 4 Estimated radicalness of major US innovations, 1953–1973. Source: US National Science Board 1977.

TABLE 6 Estimated radicalness of major US innovations by percent distribution and number of innovations, 1953–1973.

	Period			
Radicalness	1953–1973	1953–1959	1960–1966	1967–1973
	Percentage distribution			
Radical breakthrough	26	36	26	16
Major technological shift	28	17	31	35
Improvement	38	39	37	40
Imitation or no new technology	8	8	6	10
Total[a]	100	100	100	100
	Number of innovations			
Radical breakthrough	64	27	24	13
Major technological shift	70	13	29	28
Improvement	96	29	35	32
Imitation or no new technology	20	6	6	8
Total[a]	250	75	94	81

[a]Detail may not add to totals because of rounding.

SOURCE: Adapted from US National Science Board 1977.

and communications was significantly higher than that in traditional textiles or paper production. Principal technical solutions used in washing machines, refrigerators, textile machines, batteries, electric tools, combustion engines, and transport machines are, on the average, more than 25 years old, while those used in radio components, electronic calculators, and watches are generally less than 10 years old.

For a more comprehensive explanation of the productivity dilemma, we obviously must study the long-term tendencies shown by economic mechanisms and resource utilization. As we plan to investigate these tendencies in a future report, we shall not pursue the topic further here.

Over time, the productivity growth rates of various industries (see Table 4) show a developmental pattern illustrated by efficiency development in the lighting industry (see Fig. 5). The incandescent lamp, a basic innovation of the last century, reached an absolute

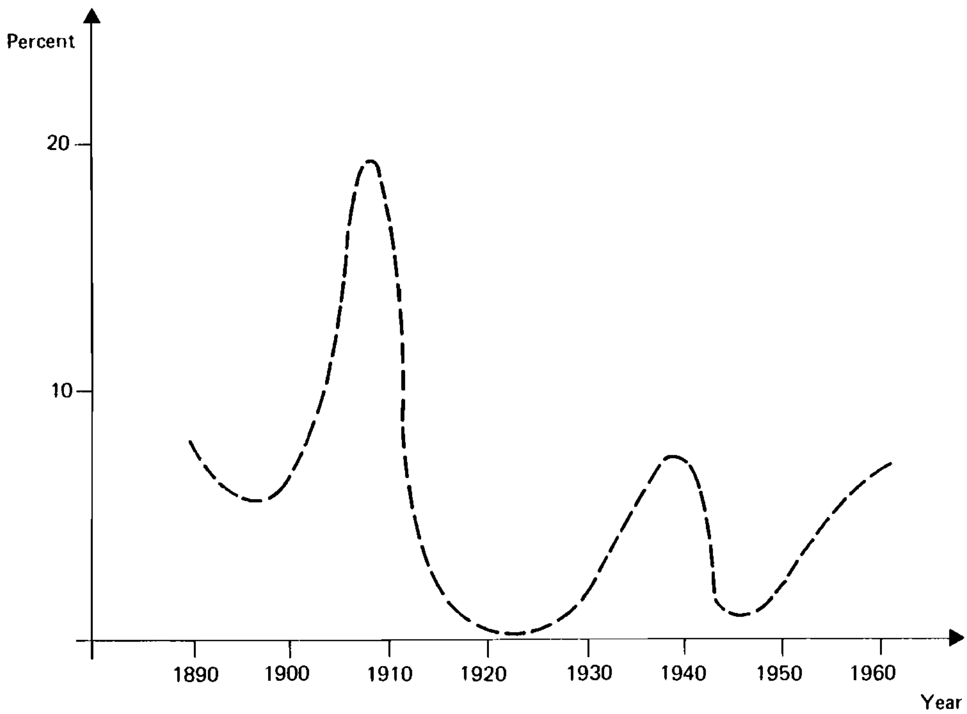


FIGURE 5 Annual percentage change in efficiency development in the lighting industry (incandescent lamps) in lmh (lumen hours of useful life) per dollar costs, 1890–1960.

peak in productivity growth rates before the First World War during a stage of rapid growth that can also be found in the developmental pattern of other industries. Such a natural trajectory is of course not only determined by the characteristics of the specific innovation process; it is also influenced strongly by the environment affecting the innovation and by interaction with other industries (see Haustein 1979). To include these factors in our consideration, we use the concept of relative efficiency, which was developed to meet the needs of planned economies (see Haustein 1976).

1.2 Relative Efficiency

The indicators of efficiency in a given production system cannot tell us whether the system is using allocated resources because of requirements imposed by the economic system as a whole. We should therefore compare these indicators with those of the next higher system (for example, a sector of industry) or with those of the entire industry.

Efficiency is the relation of output O and input J over time t:

$$e(t) = O(t)/J(t) \tag{1}$$

The efficiency of an innovating system (dynamic efficiency) is

$$e_i(t) = O_i(t)/J_i(t) \tag{2}$$

The efficiency of a higher system is

$$e_s(t) = O_s(t)/J_s(t) \tag{3}$$

The relative efficiency of an innovating system is therefore

$$e^*(t) = e_i(t)/e_s(t) \tag{4}$$

However, the efficiency of a higher system is

$$e_s(t) = \frac{\sum_{i=1}^{n} e_i(t)p_i}{\sum_{i=1}^{n} e_i(t)} \tag{5}$$

where

$e_i(t)$ is the efficiency of production system i, where $i = 1, 2, \ldots, n$

p_i is the production share of the system i

and

$$\sum_{i=1}^{n} p_i = 1 \tag{6}$$

Clearly, then, the efficiency of the next higher system depends not only on the efficiency of the innovating systems $i = 1, 2, \ldots, m$ but also on the efficiency of the noninnovating systems $m + 1, m + 2, \ldots, n$ and on the subsequent weights of those production systems. Efficiency of the innovating system that is high in comparison to that of former times may actually be a low relative efficiency if the next higher system has improved its average efficiency considerably.

Absolute or average efficiency of an innovating system is cyclical, with five stages in the cycle: take-off, rapid growth, maturation, saturation, and decline. Table 7 shows the

TABLE 7 Characteristics by stage of the efficiency cycle of an innovating system.

No.	Characteristic	Stage: Take-off	Rapid growth	Maturation	Saturation	Decline
1	Example	Solar energy	Microelectronics	Synthetic fibers	Shoemaking	Shipbuilding
2	Product-related change	Very high	High	Medium	Low	Very low
3	Process-related change	Low	Medium	High	Medium	Low
4	Number of technological opportunities	Very high	High	Medium	Low	Low
5	Dominant kind of innovation	Basic	Improvement-related	Improvement-related	Improvement-related	Pseudo
6	Dominant kind of change in production units	New establishments	Enlargements	Total modernization	Rationalization	Rationalization
7	Technological policy	Push	Push and compensation	Compensation	Compensation	Compensation
8	Patent activity	High	Very high	Medium	Low	Very low
9	Economic organization	Very flexible	Flexible; increasing number of firms	Increasing vertical integration; high economics of scale	Increasing diversification; declining number of firms	Declining number of firms
10	Competitive situation	Performance of product dominant	Performance of product dominant	Quality dominant	Price dominant	Outsider as innovator
11	Export policy	Low export activity	High share of exports	Declining share of exports	Production moved abroad	Production moved abroad
12	Labor demand	Rapidly increasing	Increasing	Static	Decreasing	Decreasing
13	Capital intensity	Low	High	High	Very high	High
14	Personnel requirements	Scientific and engineering expertise	Management skills	Unskilled and semiskilled labor	More skilled labor	Drastic reduction in employees required
15	Management	Informally organized and prone to take risks	Dominated by entrepreneurs	Dominated by experienced organizers	Bureaucratic	Marked by change in upper levels
16	Societal need	Very high	High	Medium	Medium	Low
17	Demand	Low	High	Very high	Medium	Low
18	Absolute efficiency (growth rates)	Very low	Very high	High	Medium	Low
19	Allocation of resources	Low	Medium	High	Very high	Medium
20	Total benefits	Very low	Medium	High	Very high	Low

cycle's characteristics, which we derived from case studies. Number 1 gives examples of industries in various stages, while numbers 2–8 describe technological features. The trade-offs among these indicators are significant for technological policy in an industry. For example, there is no congruence between product-related change (2) and process-related change (3), especially in the first three stages. We need to determine whether the decrease in efficiency growth rates of product-related change from take-off through decline can be compensated by the efficiency growth rates of process-related change, and if so, for how long. Numbers 9–17 describe the cycle in economic terms. Managerial requirements obviously differ over the five stages. Fluctuations in efficiency often result from managers' slow or inappropriate reaction to changes. Numbers 18–20 show a more aggregated trade-off. Growth rates of absolute efficiency (18) are normally highest during rapid growth, but the absolute sum of benefits (20) is normally highest during saturation; thus managers are often unaware of the transition threatening to lead to the last stage, decline.

Table 8 reflects the developmental patterns of leading industries in the FRG, where structural change resulted from a number of basic innovations used after the Second World War. However, we should not forget that an innovation is always the fusion of economically relevant demand and technical feasibility.

TABLE 8 Share of innovative industries (in percent) in the net production of the manufacturing and mining industries in the FRG, 1950–1977.

Industry	Share in net production of manufacturing and mining industries in the FRG (in percent)						
	1950	1955	1960	1965	1970	1975	1977
Petrochemicals	0.88	1.30	2.22	3.33	3.80	3.56	3.47
Plastics	0.22	0.40	0.73	1.20	1.73	2.34	2.57
Aircraft engineering	–	–	0.15	0.30	0.45	0.45	0.40
Chemicals	7.05	7.06	7.08	8.45	10.51	11.77	12.23
Electronics	4.84	6.84	8.19	8.93	9.96	11.06	11.72
Automobile engineering	2.94	4.53	6.04	6.64	7.49	7.32	8.13
Total	15.93	20.13	24.41	28.85	33.94	36.50	38.52

SOURCE: Adapted from Krengel *et al.* 1973, 1975, 1978.

The higher efficiency of an entire industry no doubt accounts for rapid development in the industry's innovative sectors, but data also indicate a diminishing rate of relative efficiency (see Krengel *et al.* 1973, 1975, 1978). The growth rate of labor productivity in the innovative sectors in comparison to that in manufacturing industry as a whole was significantly higher from 1950 to 1955 than from 1973 to 1977. During the 1950–1955 time span, the growth rate of labor productivity in the petrochemical industry was 2.6 percent higher; in plastics, 2.0 percent higher; in aircraft engineering, 11.4 percent higher; in chemicals, 1.4 percent higher; in electrical engineering, 1.4 percent higher; and in automobile engineering, 3.1 percent higher than in manufacturing industry as a whole. During the 1973–1977 time span, the growth rate of this factor was significantly lower: in the mineral industry, 1.9 percent lower; in plastics, 1.5 percent lower; in chemicals, 0.46 percent lower; in electrical engineering, 1.7 percent lower; and in automobile engineering, 1.6 percent lower than in manufacturing industry as a whole.

We can draw the following conclusions from these statistics and from our case studies:

1. A period of high dynamic (as opposed to average) efficiency follows the take-off stage.
2. Through better use of basic innovations the production process becomes increasingly capital-intensive and decreasingly labor-intensive. A diminishing rate of relative efficiency results, with a tendency for production units that have adopted an innovation to lose, after some time, the advantages of dynamic efficiency and to approach the average efficiency of the entire industry.
3. In the future, dynamic efficiency will depend largely on a country's ability to exploit new fields of innovation.
4. The main concern of a country in its innovation policy should be to have the optimal combination of business activities in various stages of the innovation cycle. Countries, industries, or firms concerned primarily with activities of the take-off stage may find themselves lacking sufficient economic resources to exploit these activities through improvement-related innovations. Countries, industries, or firms dominated by activities of the maturation stage, such as limitation and improvement of given technologies, incremental innovation, diversification of products, exploitation of scale economy, extension of vertical integration, and automation of production processes, will lose their advantage with respect to dynamic efficiency and experience stagnation.

To find the proper mixture of business activities in various stages of the innovation cycle, we need information about the characteristics of innovations. Distinctions that are important on the level of the production unit may be unimportant or impractical on a higher level. On the macroeconomic level, we think that it is important to distinguish between basic, improvement-related, and pseudo innovations. Basic innovations create new potential for efficiency and open new fields and directions for economic activities. Improvement-related innovations, many of which are incremental innovations, absorb this potential for efficiency by improving the given system and bringing it into balance. Improvement-related innovations become pseudo innovations at the point where they are unable to achieve higher efficiency in production.

A crucial task to improve innovation policy at the national and company levels is to provide information about future fields of innovation, which are dependent on various factors that fall into three categories:

- Urgency of demand for the innovation
- Existence of scientific and technological solutions to meet unsatisfied or latent demand
- Existence of a social environment that allows the fusion of demand-related factors and scientific–technological feasibilities

From the perspective of our current knowledge, for example, we can say that in the next two decades nations will achieve high dynamic efficiency, enabling innovation in the following fields:

- The electronics complex (especially applied microelectronics), which will make further development in automation possible

- The energy and environment complex
- Biochemistry and the food production complex
- Technologies able to provide new organizational solutions to solve communication, traffic, urban, health, and recreation problems

Successful innovators will probably be those able to respond effectively in these fields of innovation. Once the right direction is chosen, success depends on managing the factors that influence innovative activities.

2 FACTORS INFLUENCING INNOVATIVE ACTIVITIES: AN ANALYTIC APPROACH

2.1 A Model of the Innovation Process

2.1.1 Innovation vs. Invention

Innovation, a well-known term since the days of Schumpeter, should not be confused with invention (see Schumpeter 1952). Innovation includes the activities, not only of research and development, but also of technical realization and commercialization. In looking at the great number of studies and books on innovation that have been published, we noted first, the microeconomic approach used in most studies and second, the common view of innovation as a single process, a single technological change (in the narrow sense of the word *technological*). We think that innovation must be treated differently. The history of technology provides many examples where single important technical solutions had no socioeconomic impact (see Haustein 1974). We do not consider such solutions to be innovations.

The steamboat *Great Eastern*, for example, was a fundamentally new solution in the mid-nineteenth century. Its motive power was 100 times stronger than that of customary ships, while its tonnage was up to 7 times greater. Such a ship was, however, inappropriate at that time, as ports and service facilities were not able to accommodate it. After several years, the shipping trade firm that owned the steamboat, unable to withstand its economic consequences, went into bankruptcy (see Henriot 1955).

As a second example, many inventions in electrical engineering were well known a century ago. The 1883 exhibition of electrical products in Vienna included, for instance, electric water heaters, hearths, cushions, and motors, but there was no application for such devices in the existing complexes of needs and resources. Only one invention (the incandescent lamp) completely changed the existing system of demand (that for lighting). The Berlin power station was built in 1885, and until 1900 electrical demand was primarily for lighting. Electric lighting was accepted as a basic innovation for two reasons. First, a rapid increase in demand could be established in this field. Electrical illumination of the Munich opera, for instance, had a striking effect. Second, Edison, the pioneer in this area, was not only a great inventor but also a good systems engineer and entrepreneur. He built a complete system, from production and distribution to usage, for satisfying the demand for lighting. He initially set the price for one lamp at \$0.40, but costs were higher – \$1.25. After three years he was able to reduce costs to \$0.37 and to obtain large profits from an explosion in demand.

These examples suggest the difference between technological change in a narrow sense and the innovation process. Innovation always causes a change in the technological system, with a great impact on the socioeconomic system or subsystem affected. Such subsystems are

- Complexes and subcomplexes of needs or demand (e.g., demand for lighting)
- Complexes and subcomplexes of resources (e.g., sources of energy)
- Processing cycles from primary production stages to final consumption (e.g., the wood cycle from forestry to the use of furniture)

(We also differentiate between basic and improvement-related innovations from this standpoint in Section 2.5.)

2.1.2 Other Terms

After many years of conceptual confusion and dissension about the proper definition of the range of research in studies of technological innovations, we have learned that only a comprehensive and complex approach provides useful results. The need for such an approach provides us with a starting point for describing our conceptual model for analyzing the process of technological innovation. Figure 6 provides a context for the terms used. According to the procedure prevailing in innovation research, we define innovation, for the time being, as the total process of research, development, and application of a technology; this initial working definition for a limited analytical purpose omits exploitation, the fourth innovative activity shown in Fig. 6. By technology, we mean the knowledge of the properties and applicability of a technique.†

A technology may be related to a product or to a production process. Each of the innovative activities may be divided into two stages, producing the analytical sequence of innovative activities shown in Fig. 6: basic research; applied research; technological development; commercial development; application in production (of a product or of the hardware or software of a process); and application in consumption (use of a product or process).

These distinctions, which are made for analytical purposes only, are not intended to show a necessary progression over time. There may be breaks and lags, and several activities related to the same technology may be performed simultaneously. In particular, research and development – even basic research – may be carried on after a technology has been applied for many years. In pharmacy, for example, the effectiveness of new products is often recognized without certain knowledge of the way in which the products work. A product or production process long since applied may thus be the subject of investigation.

There are two ways to show the innovation process over time. For theoretical purposes, we can use a spiraling model, where time is the axis within the spiral and the spiral consists of a carousel of the six previously mentioned innovative activities. For empirical studies, however, another approach seems more adequate. The situation of a technology and an innovator (see Section 2.1.3) in the case of exploring and developing a new technology is completely different from that in the case of realizing and improving an existing, previously applied technology. Thus we supplement research, development, and application by a fourth stage, exploitation, to take into account innovative activities that may be carried on after a technology is initially applied. We chose this term to reflect the innovator's

†This distinction, which is in keeping with the historically based custom of German science, is made only for clarification; for the purpose of this report, it is sufficient to use "technology" in its usual broader sense.

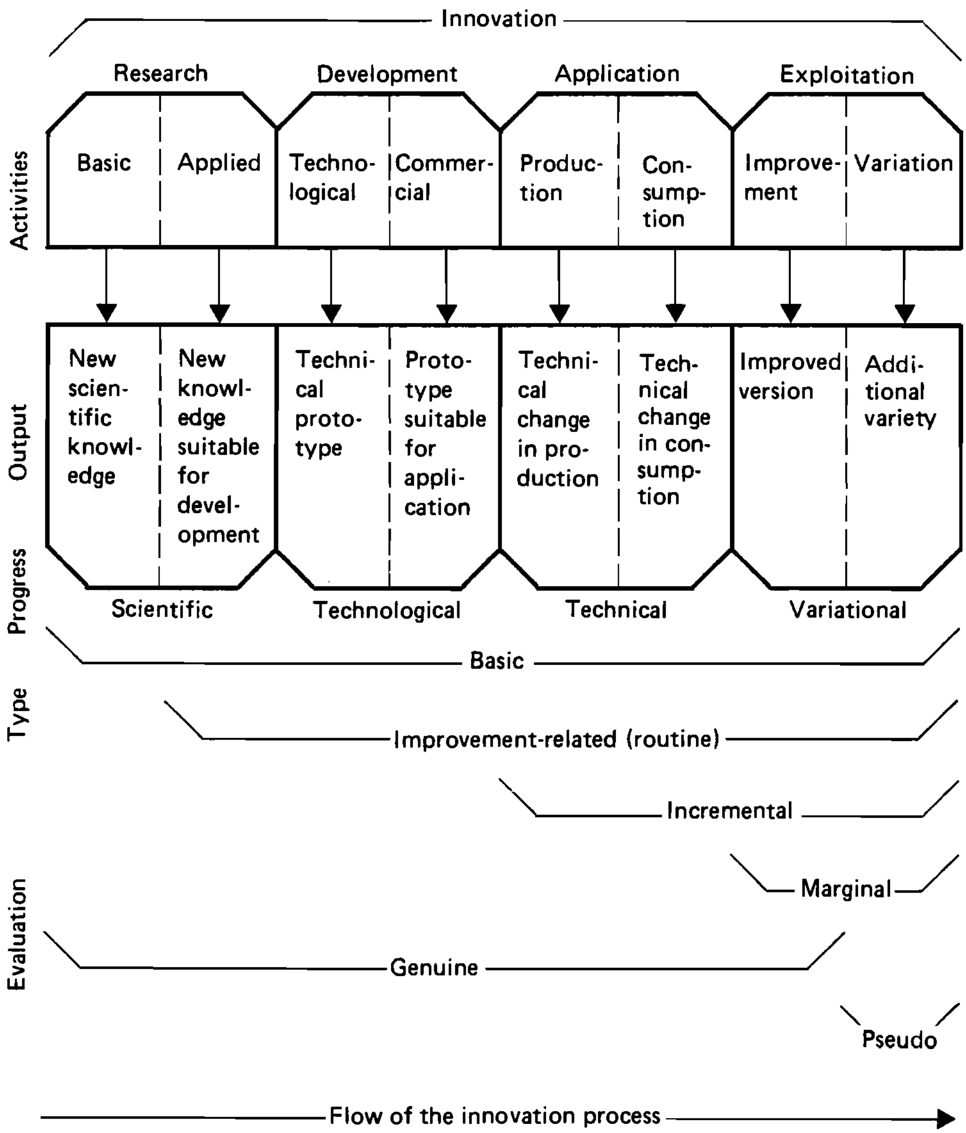


FIGURE 6 The innovation process.

propensity to make full use of the additional opportunities offered by a technology that has already been applied. The exploitation stage is also divided into two activities. improvement and variation of the technology.

Figure 6 shows that the various innovative activities result in different kinds of output: basic research, in new scientific knowledge; applied research, in new knowledge suitable for development; technological development, in a technical prototype; commercial development, in a prototype suitable for application; production, in a technical change in production; consumption, in a technical change in consumption (use); improvement, in an improved version of the technology; and variation, in additional variety.

The innovation process results in four different kinds of progress: scientific progress from research, technological progress from development, technical progress (in the technological but not necessarily in the economic sense) from application, and variational progress from exploitation. In this context, progress is a strictly conceptual term, not an assessment of the activities' results. New knowledge, a new prototype, a technical change, and additional variety of the technology are new possibilities that increase the opportunities to choose among alternatives, including those offered by existing technologies. These new possibilities are therefore kinds of progress.

We do not believe that a typology of innovation can be derived solely on the basis of a single innovation process. The decisive criterion for classifying innovations as basic, improvement-related, incremental, or marginal is related to the interaction between innovation processes and the environment. On the other hand, the major types of innovation can also be shown by their location within the scheme.

Bearing in mind that innovation research began by investigating activities related to the problems and benefits of dealing with something technologically new, we call such activities genuine innovations only when they result at least in a technically improved version of the technology under consideration. In contrast, we term activities resulting only in additional variety of the technology pseudo innovations. Genuine innovations are the real subject of innovation research. However, attempting to trace the influence of the life cycle of a technology on the efficiency of the system of which the technology is a part of course necessitates looking at the total process of innovation (i.e., the life cycle the innovation has passed through until that time), which includes activities involving both genuine and pseudo innovations.

Having dealt with the various kinds of innovative activities and types of innovation, we now turn to the innovation system and those involved in innovative activities.

2.1.3 The Innovation System

As we feel that a microeconomic approach to innovation or a definition oriented to a single process is not sufficiently operational and prefer to view innovation as a change in the technological system with a great impact on the given economic system or subsystem, we have devised a scheme with three levels representing subsystems of the innovation system. The first is the innovator, the person or group carrying on innovative activities; the second, the organization within which the innovator acts; and the third, the social, economic, and political environment of the organization. The term environment is of course general and requires explanation. In planned or market economies there is no simple "selection environment" in the biological sense of the term as used by Nelson and Winter (1977). An economic environment is hierarchically structured and consists of at least two levels, microeconomic and macroeconomic, which have their own laws and regularities. The levels must be linked, not by extending the laws of one to the other, but by studying their interaction. The economic environment surrounding innovations is an operational or policy-oriented environment that depends greatly on actions taken on the national level; this is true of both planned and market economies.

In a general sense we can define a system as a set of elements among which relationships exist. These relationships either may be of a structural nature, framing the system, or may actually take shape in the system; the latter are called process-related variables. Combining our concept of three levels with this definition of a system, we arrive at the matrix of nine cells shown in Table 9.

In economic terms, the innovation process is a production process transforming input (production factors) through innovative activities into output (progress). This concept

TABLE 9 Components of the innovation system.

	Variables related to		
Level	Elements	Structure	Process
Innovator	a. Input/output	b. Interaction among innovators	Innovative activities[a]
Organization	c. Resources	d. Organizational dimensions	e. Organizational measures
Environment	f. Resources	g. Environmental dimensions	h. Environmental measures

[a]Variables a h are factors influencing innovative activities.

connects the components of the innovation system, shown in Table 9, with the flow of the innovation process, shown in Fig. 6.

The inputs into the innovation process (such production factors as labor, capital, materials, and technological know-how) are taken from an organization's resources; the organization in turn takes and receives input from its environment. The innovator's output (the various types of progress) augments the resources of the organization and of its environment. The transformation process (the shape of the production function) is determined by the quantity and quality of these resources and by the measures (steps, actions) taken by the organization and the environment to change the organization, which influences innovative activities. Furthermore, it is determined by the dimensions (general features) of the organization and of the environment, by the interaction among innovators, and, accordingly, by the efficacy of the process. Consequently, to draw conclusions about the efficiency of innovative activities in a given context (technology, time period, area), we must determine the factors influencing the activities and their efficacy with respect to those activities.

2.2 Factors Influencing Innovative Activities

2.2.1 Groups of Factors

Research has revealed a vast number of factors affecting the innovation process, especially those acting as barriers to innovative activities. It is not feasible to compile from the literature a list of factors that simultaneously is exhaustive but does not involve overlapping or double counting of terms. Therefore, we have established from our own experience a set of factors in which we have also tried to include the results of others' work; unfortunately, it is not possible to cite all the theoretical and empirical studies, the assumptions and findings of which we have included in our discussion. Our set of factors is not restricted to those factors that empirical studies have shown to influence concrete innovations. Instead, it contains as many factors as possible that might exert an influence.

We use Table 9 as a guideline for identifying and classifying the factors (a complete list of which appears in Appendix B). Following are the groups into which they may be distributed on the three levels of the innovation system.

I. Innovator
   a. Input/output
      a1. Input-related factors (necessary quantities and qualities of factors relating to production)
      a2. Output-related factors (knowledge and utilization of the properties and possible applications of the technique)

- b. Interaction among innovators
  - b1. Interplay of functional roles (which must be fulfilled to accomplish innovative activities)
  - b2. Characteristics of innovators (persons playing these roles)

II. Organization
- c. Resources (e.g., labor)
- d. Organizational dimensions
  - d1. Relationships with the environment (e.g., recognition of clients' needs)
  - d2. Internal dimensions (e.g., system of goals)
- e. Organizational measures
  - e1. Planning measures (e.g., selection of projects)
  - e2. Control measures (e.g., supervision of innovative activities)

III. Environment
- f. Resources (e.g., capital equipment)
- g. Environmental dimensions
  - g1. Economic sector (e.g., system of competition)
  - g2. Political sector (e.g., national goals)
  - g3. Social sector (e.g., system of social values)
- h. Environmental measures
  - h1. Economic sector (e.g., cooperation with suppliers)
  - h2. Political sector (e.g., regulations)
  - h3. Social sector (e.g., public familiarity with the technology)

2.2.2 Patterns Shown by the Factors

The existence of these factors influences the performance of innovative activities; to a large extent, then, the factors govern the efficiency of innovative activities. The power of the factors to govern the efficiency of the activities (that is, their efficacy in influencing those activities) is likely to depend on certain circumstances, which we may determine by asking the following questions:

- Which factors influence which innovative activities?
- In doing so, which clearly inhibit and which clearly promote innovative activities? Which are of indistinct efficacy?
- With what strength or weight does a given factor influence innovative activities?

A given factor may, in a given situation, have the effect of a blockade, obstacle, facilitator, or incentive to innovative activities, according to a continuum of efficacy ranging from inhibiting to strongly promoting innovative activities. By combining these four possible effects with the systems approach developed thus far, we seek to gain a theoretical notion of the efficacy of the various kinds of factors before beginning empirical research, which must deal with an interwoven network of factors and activities in a particular case.

We can begin with three principles. First, we assume that the more a factor is present in a manner that is suitable (or is not present in a manner that is unsuitable) for innovative activities, the more it is likely that the factor will not stop but rather will promote these activities. In this context we shall differentiate in Section 2.3.1 among distress, slack, and excess of factors.† Second, the degree of likelihood of inhibiting or promoting innovative activities is higher on the level of the innovator than on those of the organization or environment because of the innovator's more direct and immediate influence. The farther the level is from the innovative activities, the greater is the distress to be compensated for and coped with. Third, and similarly, factors consisting of element- or process-related variables can influence innovative activities in a more direct and immediate manner than can factors related to structure. Table 10 illustrates these principles, but it can of course give only a hypothetical view of the prevailing efficacies.

TABLE 10 The prevailing efficacy of factors in the innovation system.

	Factor			
	Presence and	Related to		
Level	suitability	Elements	Structure	Process
Innovator	Distress	Blockade	Blockade Obstacle	–
	Slack	Facilitator	Facilitator	–
	Excess	Incentive	Incentive	–
Organization	Distress	Obstacle	Obstacle	Blockade Obstacle
	Slack		Facilitator	
	Excess	Facilitator	Facilitator	Incentive
Environment	Distress	Obstacle	Obstacle	Blockade Obstacle
	Slack		Facilitator	
	Excess	Facilitator	Facilitator	Incentive

Three types of change are responsible for altering the weight of a given factor during the innovation process:

A. Changes related to the stage of the innovation process
   - A1. Specific aspects of innovative activities (e.g., problems related only to research and development)
   - A2. Settlement or solution of the underlying problem (e.g., reduction of technological risk after a solution has been found)
   - A3. Shaping of the technology (e.g., insuring the success of market products)

B. Changes related to the expiration of time
   - B1. Exploitation of benefits (e.g., saturation of demand)
   - B2. Appearance of antagonists (e.g., emergence of competing firms or technologies)
   - B3. Altering of attitudes and values (e.g., boredom of those involved in innovative activities)

†Slack and distress situations were first used in innovation research by Knight (1967) in his model of the intra-firm innovation process.

C. Changes compelled by "fate"
Accidental factors that are unforeseeable – that is, not definable from within the innovation system (e.g., changes in energy prices)

Our next problem is to combine these types of change with the set of factors and to apply the result to the innovative activities in order to determine the prevailing efficacy of particular factors during the innovation process. Obviously, changes compelled by fate cannot be considered because their efficacy is not predictable.

2.3 Efficacy of Factors

2.3.1 The Concept behind Our Presentation

In examining the varying importance of the factors influencing innovative activities during the innovation process, we must concede some restrictions. First, we must concentrate on particularly striking relationships and omit those that seem to be of minor importance for the activity in question. Second, while empirical studies have provided us with a great deal of information about special features of the innovation process, no study covers all the influencing factors; for obvious reasons, no opportunity exists to carry out such a study adequately. Therefore, we are left with a mixture of evidence from empirical studies, results of theoretical reasoning, plausible arguments, and sheer truisms.

We think it best to start from the idea that innovative activities are encroachments on the existing state of life and therefore require a continuous impetus. Whether the "energy" for this impetus is provided depends on the presence and suitability of various factors. We may thus classify the factors according to the likelihood that they will act as blockades or as incentives to innovative activities, as mentioned previously. Neglecting, for the sake of brevity, the caution pointed out in Section 2.1.2 regarding the course of time and the progression of innovative activities, we may indicate the efficacy of a particular factor through the example shown in Fig. 7. The factor illustrated is presumably more likely to act as a

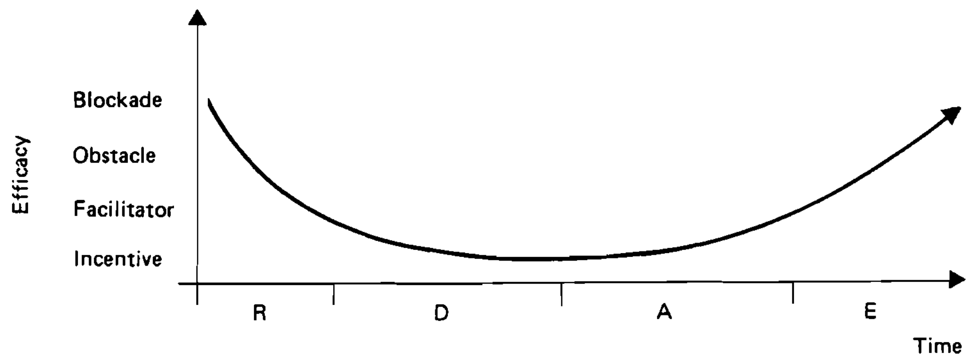


FIGURE 7 Efficacy of a factor influencing innovative activities, where R is research, D is development, A is application, and E is exploitation.

blockade early in the innovation process (overcoming this blockade would require a great deal of "energy"); then it promotes innovative activities for a time until finally blocking them once again. The factor is subject to a type of change that causes a sequence of distress-slack-excess-slack-distress, resulting in the shape of the curve shown in Fig. 7; other types of change cause other shapes of curves for various factors. Figure 8 shows the curves appearing in Appendix B, where we have also used a wavy line to indicate cases where the sequence is not predictable.

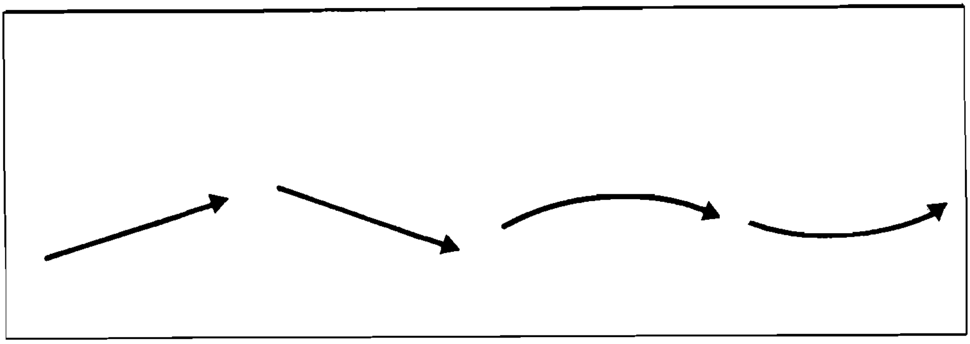


FIGURE 8 Curves representing efficacy of various factors influencing innovative activities.

We cannot determine which situation (distress, slack, excess) exists at a given level (innovator, organization, environment) of the innovation system without accurately knowing the circumstances of the subsystems at those levels. The most prudent way of tackling this problem seems to be to determine whether distress (for example) of a particular factor with regard to a given innovative activity and level of the innovation system might be a serious hindering factor. We are speculating, in other words, on the factor's efficacy in influencing innovative activities.

2.3.2 Detailed Analysis

Appendix B presents our hypotheses regarding the efficacy of factors influencing innovative activities during research (R), development (D), application (A), and exploitation (E). We treat the factors in the order presented in Section 2.2.1, give a short explanation of our hypotheses, and indicate (using the alphanumeric codes presented at the end of Section 2.2.2) the type of change most likely to predominate. Although it would be senseless to count the shapes of the various curves to find dominant characteristics, four features should not be overlooked.

First, the efficacy of most factors is determined by changes related to the stage of the innovation process: by settlement or solution of the underlying problem (A2) or, to a lesser extent, by occurrence of problems only during certain innovative activities (A1).

Second, comparison of the three levels (innovator, organization, environment) and types of variables (related to elements, structure, process) in the innovation system reveals on all three levels a succession of problems stemming from element-related variables (input/output, resources). In the early stages of the innovation process, problems may arise from a lack of adequate labor, materials, facilities, and knowledge; later on, problems may arise from capital requirements and from the impact of the technology on the natural environment. With respect to element-related variables, then, there seems to be no general tendency for increasing or decreasing efficacy.

Third, at the levels of the innovator and organization, the efficacy of structure-related variables tends to decrease over the course of innovative activities, whereas it tends to increase at the level of the environment, because the technology is increasingly implemented in the subsystems of the innovator and organization and ceases to be an extrinsic part of these subsystems. The technology may even become a part of the structure (e.g., goal system, long-term plan). At the level of the environment, however, the technology that is scaling up, requiring more and more resources, and having an increasing effect on the environment attracts more and more public attention, must overcome competition, and must be adjusted to the existing structure.

Fourth, the efficacy of process-related variables tends to decrease at the level of the organization and to increase at the level of the environment for reasons similar to those mentioned in the preceding paragraph. Measures that can be taken by the innovating organization are taken as early as possible, thus settling problems. In contrast, innovators and their organization must cope with measures stemming from the political, economic, and social sectors of their environment; these measures become increasingly relevant as the technology is exposed to the public.

Table 11 presents general conclusions drawn from analyzing factors influencing innovative activities at the level of the firm. These general tendencies are based on the hypotheses given in Appendix B and cannot, of course, be more than "macro-hypotheses"; if they are valid, the consequences are clear. One political implication, which we shall simply mention, is that because there is a sequence of tendencies related to the efficacy of factors influencing innovative activities, there is a corresponding sequence of priorities for policy-oriented measures designed to intensify incentives and to remove blockades to innovative activities; thus there must be many measures available to policy makers. The consequence of interest to us here, however, is the significance of the efficacy of various factors. The sequence of efficacies implies a sequence of incentives, facilitators, obstacles, and blockades in the innovation process.

2.4 Control of Factors

Our approach to determining the respective efficacy of various factors revealed a shifting of problems from the level of the innovator to that of the environment. This transition is easily understandable, as the purpose of any innovation process is to transfer the technology from the innovator's level to that of the environment. However, the innovator can control the factors influencing innovative activities to a much greater extent on his own level than on the level of the environment, where his ability to act on and react to factors is curtailed. Thus the likelihood that the innovator will determine the efficiency of the technology in question through purposeful methodological activities decreases.

If the current propensity is to concentrate increasingly on pseudo rather than genuine innovations, reflecting a stalemate in technology, perhaps it is because most of the factors influencing innovative activities exist on levels beyond the control of innovators. This consequence of the macro-hypotheses presented in Table 11 might explain the global decline in labor productivity: "pseudo" innovators must struggle more to increase efficiency than must "genuine" innovators, who can better overcome the factors acting as obstacles and blockades to their innovative activities. What types of innovation, then, can we distinguish from the standpoint of efficiency?

TABLE 11 Tendencies related to the efficacy of factors influencing innovative activities at the level of the firm.

Levels	Variables related to: Elements	Structure	Process	General tendencies
Innovator	a. Input/output Solving problems removes barriers related to labor, materials, facilities, and knowledge. Specific aspects of subprocesses create barriers related to capital and environment. Shaping the technology removes barriers through the technology itself.	b. Interaction among innovators Solving problems removes barriers through roles allocated and personal qualities demonstrated. Specific aspects of subprocesses govern the sequence of barriers related to role play of innovators.	–	Within the (innovator) subsystem, most barriers are removed. Barriers related to capital and environment may arise.
Organization	c. Resources Solving problems removes barriers related to labor, materials, facilities, and knowledge. Specific aspects of subprocesses create barriers related to capital and environment.	d. Organizational dimensions Many dimensions cannot be classified. Solving problems removes most barriers. Barriers may arise from implementation of the technology and from the increasing rigidity of the established structure.	e. Organizational measures Solving problems and shaping the technology removes barriers that may be relevant at the beginning of the innovation process.	Within the (organization) subsystem, barriers related to labor and resources are removed. Barriers related to capital and environment and to the rigidity of the established structure may arise.
Environment	f. Resources Solving problems removes barriers related to labor, materials, facilities, and knowledge. Specific aspects of subprocesses create barriers related to capital and environment.	g. Environmental dimensions Many dimensions cannot be classified. Specific aspects of subprocesses create barriers through public reaction, social acceptance, and the increasing relevance of the system of competitors.	h. Environmental measures Solving problems removes some barriers. Specific aspects of subprocesses and exploitation of benefits derived from the technology create barriers related to competition, social attitudes, and extension of the technology.	Within the (environment) subsystem, barriers related to competition, social values, and extension of the technology increase.
General tendencies	Resource-related problems shift from labor, materials, facilities, and knowledge to capital and environment. Shaping the technology removes technological barriers (with the exception of increasing complexity).	Structure-related problems shift from internal to external areas. Scaling up the technology attracts outside attention and widens the problem area.	Influence shifts from that of the organization to that of the environment, which may react to extension of the technology. The organization's influence is increasingly diminished as work progresses.	The development of problems governs the efficiency of innovative activities. Problems shift from the level of the innovator to that of the environment. Whereas realization removes barriers, extension creates them.

2.5 Classifying Innovations by Efficiency

There are many possible ways to classify innovations. Looking at the production process, for example, we can differentiate among innovations related to a product, to a production process, or to manufacturing. With three types of technological change (new, improved, and existing technology), we find 3^3 or 27 possible combinations. One, for example, would be a new product produced by an existing process in an improved manufacturing system. Innovations might also be classified, according to their economic results, as capital- (material-, energy-, or machine-) saving or as labor-saving.

We might also classify innovations according to

- Class of need satisfied
- Kind of resource saved
- Kind of resource processing system or industry affected
- Change in the relation between extension or rationalization investment
- Source calling for innovation
- Kind of knowledge used
- Cost involved
- Factor determining success
- Consequence
- Share of research and development needed
- Impact on the system's goals
- Component of the production process (e.g., material, machines, manpower, product, process, organization) affected
- Level of administration needed
- Size of firm involved
- Type of property used
- Degree of international competitiveness reached

Groups of interlinked innovations can be found with the help of cluster analysis; the Institute for Economic Research (IFO) study, for example, differentiated between 20 criteria and 274 features of innovation (see Uhlmann 1978). Through cluster analysis, 218 innovations were classified originally into 18 and later into the following 11 significant groups (clusters):

- Market-oriented basic innovations in large-scale organizations (enterprises)
- Cost-reducing innovations within state-owned energy-producing enterprises
- Innovations within leading noncooperative technological/industrial organizations
- Market-oriented innovations within leading cooperative private enterprises
- Cost-reducing innovations without external transfer of technology within large-scale energy-producing enterprises
- Innovations based on transfer of technology within small-scale enterprises
- Innovations based on transfer of technology within energy-distributing enterprises
- Innovations adapted by individuals
  Innovations based on trial and error
- Market-oriented basic innovations introduced according to governmental policy
- Routine innovations sponsored by multinational corporations

We do not think it is possible to construct a universal classification for innovations by using theories or empirically based methods. In establishing a system of classification, we must begin by asking, For what purpose are we doing this? We look at the innovation process from the standpoint of the national economy or its corresponding subsystems. These large systems have three goals:

- To ensure their continuing existence and function by counteracting inhibiting factors
- To ensure the balance of the system by reducing bottlenecks
- To find new ways of ensuring efficiency in a changing environment over a long period

With respect to the impact of a given technological change on a large system, we can differentiate among three functions controlling the system:

- Continuation
- Compensation
- Push

In the energy system, for example, we find the continuing use of existing primary resources. We also encounter bottlenecks in a given energy system, with increasingly negative consequences for its efficiency. It is necessary to compensate for these bottlenecks and to ensure the balance of the entire system by mobilizing new resources. We also find technological changes that not only overcome existing bottlenecks but also establish new ones. These changes act as a stimulus, pushing the existing system over a long period and thus changing it into a new one.

Table 12 shows these functions with respect to two different types of innovation. The first generally concerns giving a push to the technological level (and later, to the efficiency) of an option and often results from overcompensating for existing bottlenecks.

TABLE 12 Types of innovation and their functions.

	Function		
Type of innovation	Push	Compensation	Continuation
Basic (BI)	●●●	●	
Improvement-related (II)		●●	●●●

The second deals primarily with continuing well-known processes and compensating for bottlenecks. These two polar types of innovation, basic and improvement-related, are also known by the terminology that follows.

- Basic innovation (BI): fundamental, major, strategic, radical, or discontinuous innovation; revolutionary change
- Improvement-related innovation (II): routine, incremental, minor, tactical, rationalization, or continuous innovation; evolutionary change

2.6 The Effect of Basic and Improvement-Related Innovations on Efficiency

2.6.1 Optimization of Investments

The main function of a basic innovation is to give a push to the existing system of technology and to change it into a new system with higher efficiency. The principal function of an improvement-related innovation is to balance a given system by improving its efficiency. As basic innovations are a complex of smaller changes, in one sense the difference between the two types is relative. Basic innovations, however, consist of small changes leading over a decade or so to increasing returns, while improvement-related innovations, starting from the existing technology, lead over a similar time span of 10 years or more to diminishing returns.

The relationship between policies of push and compensation can be demonstrated through the example of investment allocation. All investments in a given industry can be subdivided into

$$I^* = I_1 + I_2 + C \tag{7}$$

where

I_1 is the investment to overcome bottlenecks with respect to technical equipment (compensation investment), per employee
I_2 is the investment to introduce new technological solutions (push investment), per employee
C is the investment for replacement (continuation investment), per employee

Optimization is necessary only for

$$I = I_1 + I_2 \tag{8}$$

The subsequent shares of compensation and push investments are

$$i_1 = I_1/I$$
$$i_2 = I_2/I \tag{9}$$

and $i_1 + i_2 = 1$.

If the main criterion for efficiency is labor productivity, we take the replacement coefficient

$$l_i = \frac{L_0 P' - L_1}{I} 100 \text{ (percent)} \tag{10}$$

where

$L_{0,1}$ is the number of employees at time 0 or 1
P' is the index of output (P_1/P_0)
I is investments

$$L_0 - L_1 \text{ is the absolute saving of labor force}$$
$$\hat{L} = L_0 P' - L_1 \text{ is the relative saving of labor force}$$

The coefficient l_i thus shows how many employees are replaced (relatively) by a given sum of investments. This coefficient differs for compensation and push investments, but in both cases we find an invariance: when investing more, replacement coefficient l_i increases up to a certain point and then decreases.

Assuming a simple dependency including this invariance, we write

$$\begin{aligned} \hat{l}_{i1} &= a_{12} i_1 - a_{13} i_1^2 \\ \hat{l}_{i2} &= a_{22} i_2 - a_{23} i_2^2 \end{aligned} \tag{11}$$

The first coefficient $\hat{l}_{i1}$ shows the relative replacement over the share of compensation investments i_1, and the second coefficient $\hat{l}_{i2}$ shows the relative replacement over the share of push investments. In general, parameters a_{ij} are different in the two cases. Compensation investments initially have rather high replacement effects, which then diminish rapidly; push investments initially have rather low replacement effects, which then increase before diminishing.

The relative economy of labor is the sum of both types of replacements.

$$\hat{L} = \hat{L}_{i1} + \hat{L}_{i2} \tag{12}$$

$$\hat{L} = I_1 \hat{l}_{i1} + I_2 \hat{l}_{i2} \tag{13}$$

$$\hat{L} = I_1 (a_{12} i_1 - a_{13} i_1^2) + I_2 (a_{22} i_2 - a_{23} i_2^2) \tag{14}$$

As $i_1 = 1 - i_2$, we find

$$\hat{L} = I[i_2(-2a_{12} + 3a_{13}) + i_2^2 (a_{12} - 3a_{13} + a_{23}) + i_2^3 (a_{13} - a_{23}) + a_{12} - a_{13}] \tag{15}$$

$$\hat{L} = I(d_2 i_2 + d_3 i_2^2 + d_4 i_2^3 + d_1) \tag{16}$$

From

$$\frac{d\hat{L}}{di_2} = I(d_2 + 2d_3 i_2 + 3d_4 i_2^2) = 0 \tag{17}$$

$$i_2^2 + \frac{2d_3}{3d_4} i_2 + \frac{d_2}{3d_4} = 0 \tag{18}$$

We obtain the optimal solution

$$i_{2(1,2)} = -\frac{d_3}{3d_4} \pm \left[\left(\frac{d_3}{3d_4}\right)^2 - \frac{d_2}{3d_4}\right]^{1/2} \tag{19}$$

Our assumption of two quadratic equations is arbitrary; it might be more appropriate to use an exponential function for this purpose. A more complicated problem is the actual statistical identification of the two types of replacement. We used data from the automobile industry in the GDR from 1955 to 1970, where motor production showed the typical behavior of compensation investments, with a low increase in equipment per employee. We compared investments of the two types, using the two interlinked subbranches (motor production and car production) of the automobile industry.

We determined the parameters in the following equations by analyzing the time series of investments and replacements of labor:

$$\hat{l}_{i1} = 25.0 i_1 - 52.3 i_1^{\,2}$$

$$\hat{l}_{i2} = 61.2 i_2 - 72.9 i_2^{\,2}$$

The absolute economy of labor for the 1955–1970 period was

$$\hat{l} = I(106.9 i_2 - 70.7 i_2^{\,2} - 20.6 i_2^{\,3} - 27.3$$

The relative economy of labor was

$$\hat{l} = 106.9 i_2 - 70.7 i_2^{\,2} - 20.6 i_2^{\,3} - 27.3$$

In

$$i_{2(1,2)} = -\frac{70.7}{61.8} + \left(\frac{70.7}{61.8}\right)^2 + \frac{106.9}{61.8}$$

we find an optimal i_2 of nearly 60 percent. Then the optimal replacement is

$$\hat{l} = 6.86 \text{ (relative coefficient)}$$

$$\hat{L} = 126{,}000 \text{ employees}$$

The real economy of labor was $l = 5.36$ and $L = 96{,}000$ employees, showing a difference from the optimal solution of 30,000 employees. The share of push investments was actually 33 percent. Of course, estimating investment allocation in the automobile industry is not simply a question of determining the share of push investments by one criterion. Our example merely illustrates the opportunities offered by modeling.

In general, we assume the efficiency of policies of push and compensation shown in Fig. 9. Although given for only one point in time, the figures shown in Table 13 for the energy field reflect the same general pattern (see also Ray 1979).

For short-term planning we prefer a policy of compensation; only for a longer perspective do we choose a policy combining push and compensation. In practice, many basic innovations dominate the efficiency of the entire system only 10 years or more after the first commercial use (Gold 1975). The primary problem is therefore the length of the optimization period. The shorter this period, the more important a policy of pure improvement becomes. The first long-term plan of a national economy oriented toward a basic innovation (electricity) – the so-called GOELRO-plan in the USSR – had a time frame of 10 to 15 years (1920–1935).

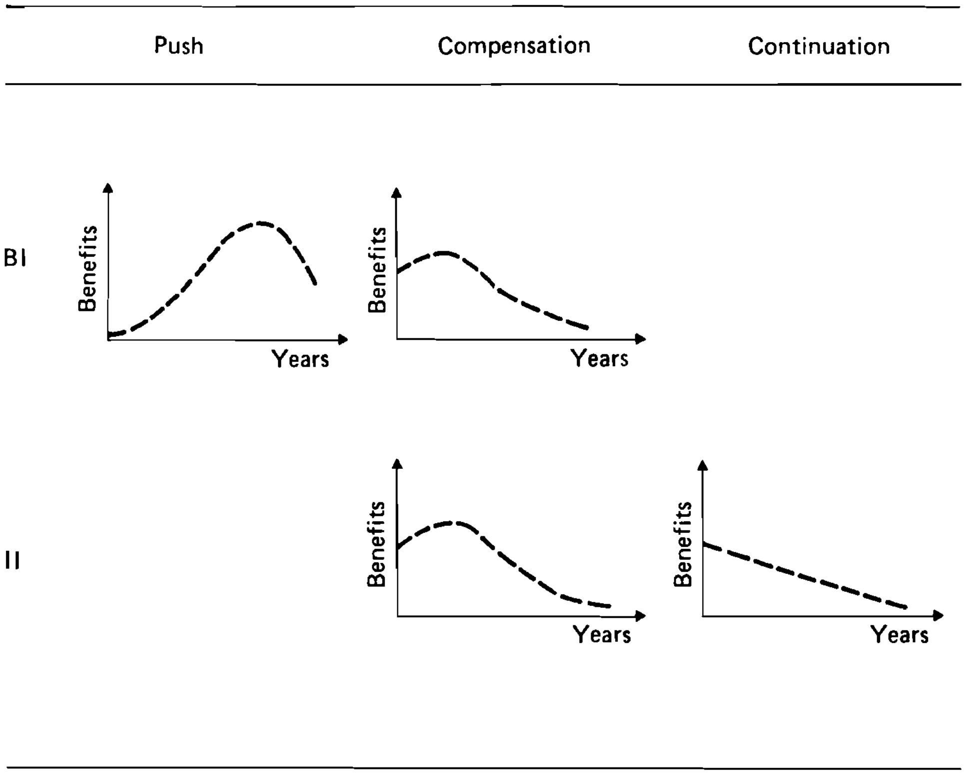


FIGURE 9 Typical progression of benefits over time under three investment policies for basic and improvement-related innovations (BI and II, respectively).

The distinction between BI and II, first made by historians (Zvorykin *et al.* 1962), was a qualitative theoretical approach. We give the terms BI and II (or the revolutionary and evolutionary technological changes cited by Nick (1974)) another interpretation. In many studies the distinction means only a certain degree of technological change. Our starting point is the influence of a given technological change on the socioeconomic system. In any given system, we find a tendency for the average efficiency to stagnate or to decrease. This tendency can be reduced by improvement-related innovations but overcome only by basic innovations whose efficiency is higher than average and whose share in output is sufficient.

While the effects of basic innovations take longer to occur than do those of improvement-related innovations, they are higher. Of course, this does not mean that we can ignore the effects of II, which are comparable over a long period to those of BI. BI and II are two sides of one coin, and the development of metallurgy proves that underestimation of II is as dangerous as fear of BI. Nevertheless, II is not able to ensure the endless efficiency of a large system. Limitless asymptotic increase of efficiency through better balancing of elements is conceivable only for a closed system. When we consider the relations of a large system with the environment, we must take into account the possibility of sudden or tremendous changes, which may lead to major bottlenecks, resource deficits, and conflict situations that can be mastered only through complex, radical solutions.

TABLE 13 The strategy of technological introduction in the energy field in the US.

Impact	Strategy	Technology	Impact in year 2000[a]
Near-term (present to 1985 and beyond)	Increase efficiency of energy use	Conservation in buildings and consumer products	7.1
	Convert waste to energy	Efficiency in industrial energy use	8.0
		Efficiency in transportation	9.0
		Conversion of waste materials to energy	4.9
	Preserve supplies of oil, gas, and coal	Direct utilization of coal	24.5
	Expand oil, gas, coal, and nuclear options	Nuclear reactors	28.0
		Enhanced recovery of oil and gas	13.6
Mid-term (1985 to 2000 and beyond)	Accelerate development of synthetic fuels from coal and shale	Production of gaseous and liquid fuels from coal	14.0
		Oil shale	7.3
	Increase use of underused fuel forms (those with limited application)	Geothermal energy	3.1–5.6
		Solar heating and cooling	5.9
	Attract more usable energy from waste heat	Utilization of waste heat	4.9
Long-term (2000 and beyond)	Develop the technologies necessary to use essentially inexhaustible fuel resources	Breeder reactors	3.1
		Fusion	–
	Develop the technologies necessary to change existing distribution systems to accommodate the distribution of new energy sources	Solar electric power	2.1–4.2
		Efficiency in electric conversion	2.6
		Transmission and distribution of electric power	1.4
		Electric transport	1.3
		Electric storage	–
		Hydrogen in energy supplies	–
		Fuels from biomass	1.4

[a]Quads = 10^{13} Btu (British thermal unit).

SOURCE: Adapted from US Energy Research and Development Administration 1976.

As a result of delay in their realization, basic innovations may have a compensatory function without stimulating efficiency during the first step of application. The energy study conducted by Häfele at IIASA showed that in using final energy we can expect many improvement-related innovations (Energy Systems Program Group of the International Institute for Applied Systems Analysis 1981). This helps us to reduce the primary energy/GDP coefficient in developed countries from the present value of 0.8 to 0.5 and in less-developed countries from 1.5 to 1.0 (Maier 1979). Conversely, the same study indicates that we must be aware of a completely different development with respect to such basic innovations as nuclear energy, synthetic fuels, solar energy, and biogas. In the next two decades, we expect a rising primary energy/GDP coefficient resulting from extensive demand pull and from delay in mastering the economy of basic innovations (see Mensch 1976).

2.6.2 Potential and Actual Outcomes

We have mentioned only the functions of innovations that contribute to achieving the goals of large systems. However, some innovations that seem appropriate for meeting the goals of a socioeconomic system or subsystem actually have a generally negative influence on it over a long period. We call such an innovation, the primary or secondary consequences of which damage the system's efficiency, a pseudo innovation (PI). We find many pseudo innovations in the consumer goods industry. In American supermarkets, where about 1500 new products appear each year, less than 20 percent survive more than one year on the shelves; the rest have proved unsellable, faddish, risky, or unprofitable or have been made obsolete by competitors with other new products. Furthermore, positive technological changes with positive socioeconomic potential can appear as innovations that have negative effects.† As Table 14 shows, a major technological change (potential BI) may thus occur only as an II or as a PI. The actual outcome depends on the ability to use innovative potential by changing many conditions necessary for optimal efficiency of the new or renewed system. As all these conditions change over time, a potential BI may or may not become an actual BI. For example, automation of the production process in a given (nonautomated) industry is a BI. It may become an II if changing the traditional process is not possible, but such automation without process-related change is not efficient. It may also become a PI; solar energy, for example, is a potential BI that may actually occur only as a PI – as in cases where solar heating systems are installed in existing buildings without changing other conditions. Similarly, an innovation planned as an II might actually function as a BI; we often do not clearly realize the qualitative or quantitative potential of an innovation. A PI might become an II as a result of learning induced by negative results.

As many innovations are closely linked over time, it is important to realize and to promote positive feedbacks in the innovation process. For example, the introduction of the railway system led to higher coal demand, and higher coal demand required better transport, which was possible through the railways. The prehistory and history of basic innovations are made up of groups of small innovations. The incandescent lamp, for example, was a BI in which many small changes were needed, and from Edison's time on, its development has been a complex of improvement-related innovations. We can differentiate

†We refer again to the distinction, made in Section 2.1.1, between innovation and invention.

TABLE 14 Examples of potential and actual outcomes of basic, improvement-related, and pseudo innovations (BI, II, and PI, respectively).

	Actual outcome		
Potential outcome	BI	II	PI
BI	Automation in connection with new production processes	Automation without changing the established production process	Retrofitting residential buildings with solar heating systems
II	Oxygen process in metallurgy	Improved performance characteristics of machines	Higher speed and motive power of automobiles
PI	Does not occur in reality	Change in advertising made for the benefit of the manufacturer but eventually useful to the consumer	Change in product with no real effect on the consumer

between improvement-related innovations leading to basic innovations and improvement-related innovations using the efficiency potential of basic innovations. BI is the result of a long process of selection in a wide field of smaller innovations that are competing with each other; it is essentially a package of technological changes creating a new system. A new BI establishes a greater potential for efficiency that can be more or less fully mobilized only through many improvement-related innovations. We call this incremental innovation.

2.6.3 A More Detailed Approach to Classification

The technological level, range of application, and impact on the national economy of basic innovations differ greatly. The technological level is closely connected with the necessary type and amount of mission-oriented fundamental research, applied research, and development, so it is understandable that the authors of the IFO study proposed to call basic innovations all technological changes that go through research and development stages (Uhlmann 1978). Another extreme is to use the term only for the main historical breakthroughs in technology, such as the steam engine, tool machine, and electricity. We cannot call pure scientific or technical results (inventions) basic innovations, as they are only first steps; their eventual classification depends on the availability of resources, socioeconomic needs, and capability of a given society for mastering the inventions. Thus it is not possible to speak about BI without considering social factors.

We propose calling basic innovations major technological changes that

- Are based on fundamental and applied research
- Have a well-defined high range of application – that is, modify essentially the existing demand or application complex (e.g., synthetic fibers), establish a new demand or application complex (e.g., television), or change the entire system of needs (e.g., production and consumption of electricity)
- Are connected with new scientific/technological principles of a higher order

BI greatly stimulates the entire socioeconomic system, has an enormous potential for efficiency, and is able to arrest or alter the tendency to decreasing efficiency in using resources.

The technological level of innovations is also an important indicator, but its connection with the efficiency of the system affected is not linear. Some basic innovations of the past, such as Hargreaves' machine, were not based on new scientific/technological principles. On the other hand, some innovations of a high scientific/technological level, such as the coal arc lamp of the nineteenth century, have not found a wide range or field of application.

Tables 15 and 16 illustrate various kinds of BI and II. We can also distinguish among three kinds of PI:

PI1 Simple product-related innovations that do not improve the efficiency of the user's system (e.g., many modifications in automobiles)

PI2 Innovations that improve the efficiency of one process but reduce the efficiency of the system as a whole (e.g., plastic materials that are inappropriate for practical needs)

PI3 Innovations that improve the system's efficiency in the short term but eventually lead to large losses or imbalances (e.g., process-related innovations in the chemical industry that later have a negative influence on the environment)

TABLE 15 Description and examples of three kinds of basic innovation (BI).

No.	Type of innovation	Share of fundamental research	Share of applied research	Range of application	Impact on production system	Lag between invention and large-scale application	Example
BI1	Major	High	High	Change in entire system of needs	Change in entire production system	20 to 60 years	First industrial revolution
BI2	Middle	Middle	High	Establishment of new demand complex (or market)	Creation of industrial branches	20 to 30 years	Microelectronics
BI3	Minor	Low	Middle	Essential modification of existing demand complex	Creation of new lines of industry	10 to 20 years	Synthetic fibers

TABLE 16 Description and examples of four kinds of improvement-related innovation (II).

No.	Type of innovation	Share of fundamental and applied research	Share of development	Range of application	Impact on production system	Example
II1	Very important	Middle	High	Establishment of new demand complex (or market); new product in existing demand complex	Creation of new industrial subbranches	Polyester
II2	Important	Low	Middle	Essential modification of existing demand complex; new parameters of well-known products	Creation of new product lines or processes	Thomas-gilchrist process Electric toothbrush
II3	Normal (incremental)	None	Low or none	Simple modification of existing demand complex; improved parameters of well-known products	Improvement in product lines or processes	Fluoride toothpaste
II4	Evolutionary change (marginal)	None	None	Little improvement	Little improvement	Better "touch" on telephones

Classification of three kinds of BI, four kinds of II, and three kinds of PI gives us the following ten kinds of innovation (I1–I10):

				I					
	BI			II				PI	
BI1	BI2	BI3	II1	II2	II3	II4	PI1	PI2	PI3
I1	I2	I3	I4	I5	I6	I7	I8	I9	I10

Looking at the ocean of innovations of course reveals a continuum not measurable by one clear indicator. Rather than considering this only as a continuum, however, we must take into account the obvious turning or break-even points in complexity, efficiency, and manageability in the total field of innovation. For instance, in socialist countries each scientific/technological task of one planning cycle is associated with one level of administration, from the firm to the national economy. Each type of task has various prerequisites in management and planning.

These are the most important relationships from our viewpoint; we do not want a complete or eclectic classification of all kinds of innovation. Instead, we concentrate on the process of transition from a given structure of technologies to a new structure that is able to overcome socioeconomic bottlenecks and major gaps in resource processing systems. Table 17 shows a more sophisticated classification by technological level and range of application that enables us to differentiate among 49 kinds of innovation.

2.6.4 An Innovation Level Index

The next step in establishing an innovation classification could be a quantitative evaluation by a technology level index. This step was made in an OECD investigation of 1246 innovations in five countries from 1953 to 1973 (see Table 18). While the linear level index used by the OECD study is given in column (1) of Table 18, we think that an exponential level index (column (2)) is more appropriate because the distance between basic and improvement-related innovations should be greater than the distance between different kinds of improvement-related innovations. The frequency distribution in column (4) also points to an exponential pattern. Another argument is the exponential growth of technological parameters during the transition to new principal solutions and the exponential saturation in the period of improvement. If we assume that the importance of innovations w (a coefficient between 1 and 100) follows an exponential function and the two parameters i_k and v_k are connected in a multiplicative form, we can write

$$w = i_k v_k \tag{20}$$

$$w = e^{ak} e^{bk} \tag{21}$$

$$w = e^{(a+b)k}$$

Taking a simple symmetrical scheme ($a = b$), we then have

$$w = e^{2ak} \qquad k = 0, 1, \ldots, 6$$

TABLE 17 Classification of innovations by scientific/technological level and range of application.

			Range of application						
			II				BI		
No.	Scientific/ technological level	ν_k	Quantitative growth of existing demand	Simple modification of existing demand complex (improved parameters of existing products or processes)	Essential modification of existing demand complex (new parameters of existing products and processes)	Development of new demand (new product or process) in existing demand complex	Essential modification of existing demand complex (new products or processes)	Development of new demand complex or subcomplex	Change in entire system of needs
	i_k		1	1.5	2.2	3.2	4.6	6.8	10
1	Quantitative growth of existing technical basis	1	1	1.5	2.2	3.2	4.6	6.8	10
2	Improvement within well-known technical principle	1.5	1.5	2.3	3.5 (Bentwood furniture)	4.8 (Bicycle)	6.9	10	15
3	Improvement within well-known technical principle with essential changes in one factor (materials, tools, or function design)	2.2	2.2	3.3 (Oxygen process)	4.8 (Thomas-gilchrist process)	7 (Diesel engine)	10 (Paper production)	15	22
4	Improvement within well-known technical principle with essential changes in several factors	3.2	3.2	4.8	7 (Stitching bond)	10 (Atomic ice-breakers)	15 (Electric railway)	22	33 (Spinning jenny)
5	New solutions within well-known basic principle	4.6	4.6	6.9	10 (Gyrocompass)	15 (Polyethylene)	22 (Detergents)	33 (Vacuum lamp)	46
6	New basic principle within same form or structural level of matter	6.8	6.8	10	15	22	33 (Synthetic fibers)	46 (Incandescent lamp)	68
7	New basic principle changing form or structural level of matter	10	10	15	22	33	46 (Radar)	68 (Transistor)	100 (Electricity)

NOTE: Examples are given for illustrative purposes in some cases.

TABLE 18 Level and frequency of innovative activities in five OECD countries, 1953–1973.

Type	Linear level (0–100) (1)	Exponential level (1–100) (2)	Frequency (absolute) (3)	Frequency (percent) (4)
Marginal	0–44	1–2	760	61
Normal II	45–55	3–5	239	19
Important II	56–66	6–10	149	12
Very important II	67–78	11–21	62	5
Radical II	79–89	22–46	29	2
BI	90–100	47–100	7	1
	0–100	1–100	1246	100

SOURCE for columns (1), (3), and (4): OECD Study as cited in Mensch 1976.

According to $1 \leqslant w \leqslant 100$ (percent), we find for $k = 6$

$$100 = e^{12a}$$

$$a = \frac{\ln 100}{12} = 0.38376$$

From this we find the coefficients of importance for each level within the $7 \times 7 = 49$ field (see Table 17).

When we try to adjoin one innovation to the $7 \times 7 = 49$ field, we realize that we often have difficulty in making an exact estimation; we thus find it inappropriate to make the classification too sophisticated. This does not mean that for special studies and innovations we do not need a more detailed typology.

Stages, classes, and types derived from a single innovation process are an important analytical tool, but they are not so useful in studying the behavior of industrial organizations. We must relate the innovation process to the activities and life cycles of industrial organizations and examine the process in relation to the growth of efficiency in industries, corporations, and enterprises.

3 INDUSTRIAL ORGANIZATIONS AND EFFICIENCY: A SYNTHETIC APPROACH

3.1 Innovative Activities in the Life Cycles of Industrial Organizations

Firms that have been successful for decades may suddenly fall back in their economic performance because of stagnation in an entire branch of the industry, the cumulative effects of years of mismanagement, or insufficient adaptation to market changes. In these situations measures that once had positive results often complete the disaster. For example, diversification sometimes is a profitable strategy and sometimes produces failure. Often an innovation itself becomes a failure, as in the case of the video disc, a record for video reproduction.

Why might the same factors or determinants have different consequences? In our opinion, the main reason lies in the trade-off between the discontinuous pattern of technological progress and the continuous pattern of human learning, or – put more broadly – in the incongruity and contradictions between technological and social progress (see Goldberg 1980).

From the standpoint of an industrial organization, the innovation process includes the life cycles of:

- The generic product of the broader, defined area of activity (including a sum of shorter life cycles of single products)
- The generic process (including a sum of shorter life cycles of single processes)
- An industry (including a sum of shorter life cycles of production units)
- The management in one industrial organization (including managerial organization and the qualifications and capabilities of managerial personnel)

The interaction of these life cycles against the background of the entire business and the national economy produces the efficiency cycle.

Observation of the life cycles of products and processes can be important for the allocation of resources in research and development. Patent applications are good indicators; they can be an early warning, at least in the case of investment policy. For example, a stitching bond (Nähwirken) invented in the GDR after the Second World War was an enormously productive textile process. Figure 10 shows that the number of applications for patents on the stitching bond in the US, GDR, UK, FRG, Czechoslovakia, and France

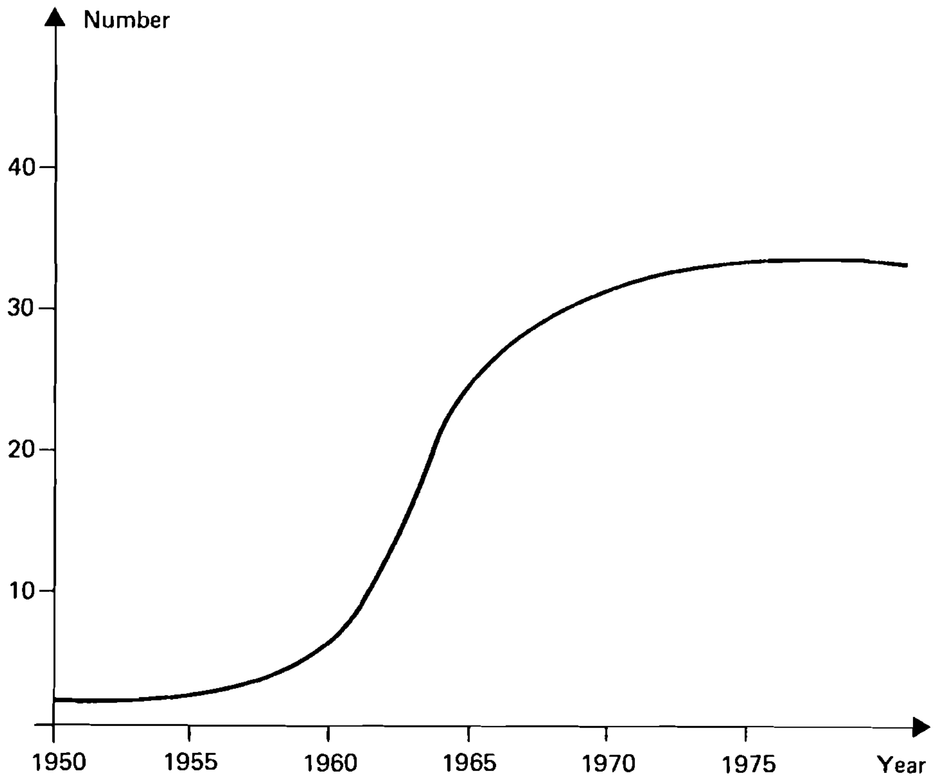


FIGURE 10 Number of applications for patents on the stitching bond in the US, GDR, UK, FRG, Czechoslovakia, and France.

was as follows: 1950–55, 1; 1956–60, 4; 1961–65, 20; 1966–70, 31; and 1971–75, 32. Production of the stitching bond, which is now used in 35 countries, is still in the stage of rapid growth, showing that saturation in patent applications is reached long before saturation in production growth.

The life cycles of products and processes are primarily technological life cycles, while those of industries and management are more complex and socially determined. Life cycles of products and processes are well-known phenomena; Abernathy (1978) and Abernathy and Utterback (1978) analyzed the interrelationship of these two cycles in the automobile industry.

Another life cycle also found in industry is more complex and includes changes in technology, organization, and qualifications of personnel. This is the life cycle of the entire manufacturing process in a given production unit (see Fig. 11). The modernization cycle

S_{00}	Starting up of production
J_{00}	First improvement stage
E_{00}	First enlargement
J_{01}	Second improvement stage
L_{00}	First local modernization
J_{02}	Third improvement stage
•	
•	
•	
M_{11}	First total modernization
J_{11}	First improvement stage after total modernization
L_{11}	First local modernization after total modernization
•	
•	
•	
M_{21}	Second total modernization
•	
•	
•	

Time

FIGURE 11 Life cycle of a production unit.

is the time between two total modernizations of the entire production unit. In the textile industry of the GDR, this has been approximately 25 years and is now approximately 18 years on the average; it may be shorter or longer in other industries.

The problem for management on the corporate level is to synchronize the individual modernization cycles of production units, including shutdowns and new establishments, to the product and process cycles and to the human factor, including changes in managerial organization and in the qualifications and capabilities of managerial personnel. The life cycle of management for European industrial organizations may be different from that in the United States. We often find a certain type of conservative manager in traditional branches and more dynamic people in more dynamic branches. When sudden changes occur in traditional industries, we can expect a complicated process of adaptation.

In analyzing the factors affecting innovations in order to grasp the human factor, we found four determinants of success or failure: innovative potential, strategic orientation, capacity for mastering ongoing processes, and cooperation and coordination. These are of varying importance in research and development, production, and marketing. The human capacity is at the same time a social capacity. Social organization and learning can change the pattern of efficiency, which is originally determined by technological progress; as yet, however, no studies have dealt with the interference between technological cycles and cycles of organization, qualification, and management.

The experience of Ericsson, a Swedish company, provides an example of the decisiveness of the human and social factor. This firm changed successfully from electromechanical to electronic telephone exchanges. It was able to do so because its managers succeeded in persuading hundreds of department chiefs to abandon their traditional working procedures and to begin a new experience. On the other hand, the conservative business ideology evident in the saturation stage of an industry creates barriers to innovation at this stage. It is responsible for a growing insistence on short-term efficiency, the "not invented here" syndrome, the formalization of short-term activities that discourage longer-term innovation projects, and a preference for a policy of compensation.

Newly established technological systems bring about new kinds of imbalance and great opportunities for such a policy. From a short-term perspective, a policy of compensation offers more benefits than does a push policy, but it can undermine the development of new possibilities. We must therefore stress again our theoretical concept of dynamic efficiency and stability. Dynamic efficiency, a kind of relative efficiency, is the real efficiency of a production system in relation to a normative efficiency or to the average efficiency of the industry as a whole. Therefore, efficiency of a production unit is a function not only of the particular cycle but also of the industrial cycle as a whole. Dynamic stability is derived from dynamic efficiency, which can be ensured only by a trade-off between push processes, which change the production system, and those of compensation, which improve it. We identify this trade-off as dynamic stability. Relative efficiency develops over the four stages in the shape shown in Fig. 12. In the maturation stage benefits are the highest in absolute measure. We cannot judge only from the standpoint of relative measures; a high profit rate may be nothing if it refers to negligible outputs.

3.2 Determinants of Innovative Activities in Industrial Organizations

3.2.1 Factor Analysis

Factor analysis of innovations can be made for various purposes. Many such analyses exist in the literature, as, for example, the study by Myers and Marquis (1969), the report on project SAPPHO (Science Policy Research Unit of the University of Sussex 1972), "The Flow of the Industrial Innovation Process" among the 218 cases cited by Uhlmann (1978), and others. The Myers/Marquis study provided an overview of factors affecting innovations and their proportions in various branches. Project SAPPHO compared pairs of successful and unsuccessful innovations, with statistical results indicating that innovations that had achieved commercial success could be distinguished from those that had failed by superior performance in five major areas:

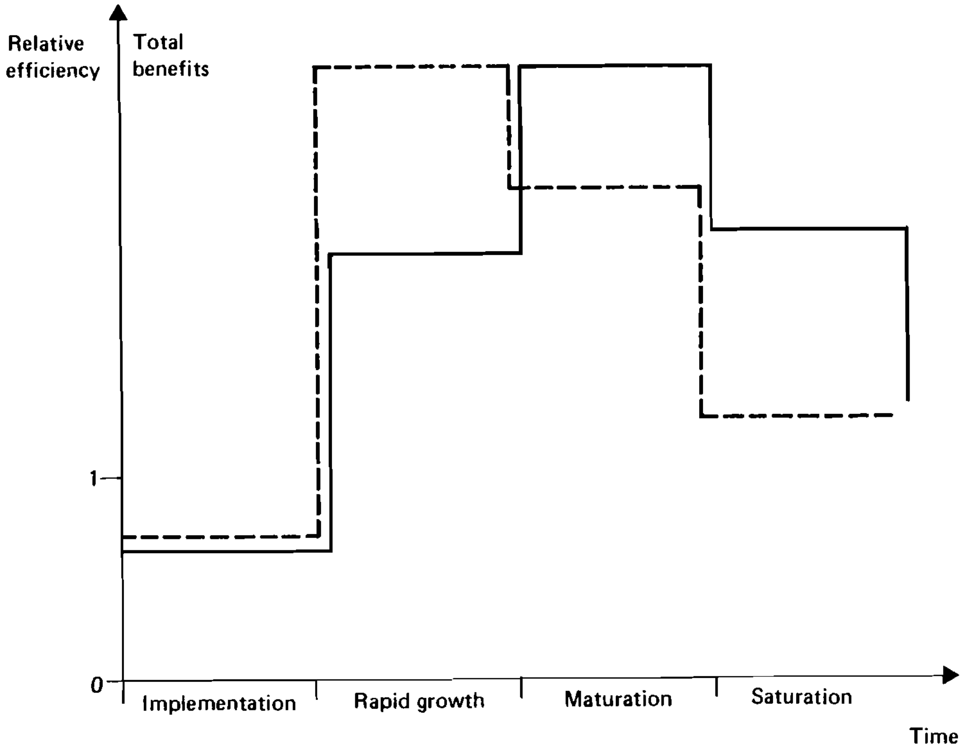


FIGURE 12 The development of relative efficiency over the stages of the innovation process, with efficiency coefficient *e* (– – –) and benefits total *E* (——).

- Strength of management and characteristics of managers
- Understanding of users' needs
- Marketing and sales performance
- Efficiency of development
- Effectiveness of communications

The Uhlmann study attempted to identify types of innovations that could be distinguished from each other by various kinds of factor combinations. All these studies were intended to serve the specific purposes of market economies, but they included not only market activities of corporations and enterprises, but also the impact of governmental policy on innovation.

3.2.2 Our Investigation of 32 Firms and Its Results

Central management and planning play an important role in planned economies, but we cannot ignore the activity of enterprises with respect to the market. We wanted to answer two questions through factor analysis: How strong is the influence of factors inhibiting the innovation process on the level of state-owned enterprises? And how strong is the influence of a firm's own ideas and measures in overcoming bottlenecks in and barriers to the innovation process?

We formulated the 26 variables shown in Table 19. We then questioned managers from 15 state-owned enterprises, using a list initially consisting of 20 and eventually including these 26 variables. We randomly chose 32 successful innovations (9 products, 9 production processes, 7 materials, and 7 manufacturing processes) in 32 enterprises, and asked the managers responsible for these enterprises the following questions:

- What inhibiting intensity *p* did the 26 variables have for the innovation concerned?
- What promoting intensity *q* did the firm's own measures have for these variables?

We asked them to rate the degree of influence according to a scale of 0 for no importance, 1 for little importance, 2 for medium importance, 3 for great importance, and 4 for very great importance. Our aim was to identify the firm's capacity to overcome barriers to and bottlenecks in the innovation process. We expected some correlation between the inhibiting intensity of the variables and the promoting intensity of the firm's activities.

The correlation coefficient between *p* and *q* was 68.82 percent over 32 innovations and 79.22 percent over 26 variables. Both are statistically significant at an error level of less than 0.1 percent. We needed to investigate more closely the specific patterns of influence for certain combinations of variables. Table 19 shows the number of statistically significant correlations between the variables. According to this and to the average values of *p* and *q* we obtained the results shown in Table 20.

The five most important variables inhibiting innovation in the 32 firms were

- Inability to master the process after release by the development group (6)
- Insufficient supply of machines and means of rationalization (4)
- Differences of opinion between managers and experts (10)
- Developmental failures not abandoned (5)
- Failures of management; insufficient interest on the part of managers (8)

The five most frequently interlinked variables inhibiting innovation were

- Differences of opinion between managers and experts (10)
- Conservative and obsolete views (15)
- Uncoordinated development among several branches (24)
- New solutions replacing the initial project (26)
- Changing demand (18)

The five most important variables promoting innovation were

- Better coordination with management (9)
- Own production of means of rationalization (4)
- Reduction of failures in developmental stages (5)
- Improvement in management (8)
- Improvement in technological and qualitative level (14)

The five most frequently interlinked variables promoting innovation were

- Better transfer of know-how between branches (20)

TABLE 19 Number of statistically significant correlation coefficients between 26 variables influencing innovations for inhibiting intensity p and promoting intensity q.

No.	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	Insufficient supply from the supplier industry	—		q									p, q				p, q	p, q	p, q		q						
2	Technical difficulties	p	—						q		p, q					q											
3	Stress caused by other production tasks			—						q	p, q						q	q	q	p, q	q		p, q				
4	Insufficient supply of machines and means of rationalization				—		q							q								q					
5	Developmental failures not abandoned					—				q																	
6	Inability to master the process after release by the development group						—																				
7	Lack of personnel in research and development							—			p, q									p, q				q			
8	Failures of management; insufficient interest on the part of managers					p			—												q	q	q				
9	Long time required by managers for coordination									—												q					
10	Differences of opinion between managers and experts		p	p	p	p		p	p		—						q			p, q				q			
11	Failures in preparation for production							p				—						q	q	q	q						
12	Delay in construction activities	p											—					p, q			q						
13	High costs; planned economy not reached													—		q									q		
14	Insufficient technological and qualitative level														—	q									p, q		
15	Conservative and obsolete views							p			p	p	p			—		q	p, q	p, q	q		q		p, q		q
16	Inexact and changing objectives	p															—	q	p, q	q							
17	Delay in recognition of problems; failures in communication	p					p						p					—	p, q	q	q	p, q					p
18	Changing demand	p									p		p			p	p	p	—		q						
19	State orders limiting the project			p				p			p					p				—	q		q		q		
20	Insufficient transfer of know-how between branches						p														—	q	q				q
21	Economizing measures																	p				—					
22	Unfavorable price relations			p																			—				
23	Insufficient special knowledge		p													p								—			
24	Uncoordinated development among several branches				p			p			p		p		p	p					p				—	q	q
25	Better solutions from competitors							p	p			p				p				p					p	—	q
26	New solutions replacing the initial project	p		p	p						p		p						p						p	p	—
	Number of significant relations for p	6	3	4	3	2	2	6	3	0	11	3	6	0	1	9	2	5	7	5	2	1	1	1	9	7	8
	Number of significant relations for q	6	3	8	3	1	1	3	4	3	6	4	3	3	2	10	7	11	7	10	12	5	8	3	7	3	6
	Sum of significant relations for p and q	12	6	12	6	3	3	9	7	3	17	7	9	3	3	19	9	16	14	15	14	6	9	4	16	10	14

NOTES:
◢ Correlation for p.
◤ Correlation for q.
■ Correlation for both p and q.
Level of significance: 0.05.

TABLE 20 Variables influencing innovations in 32 enterprises, with their inhibiting intensity p and promoting intensity q.

No.	Variable	Σp	Σq
1	Insufficient supply from the supplier industry	19	18
2	Technical difficulties	10	6
3	Stress caused by other production tasks	12	9
4	Insufficient supply of machines and means of rationalization	2	2.5
5	Developmental failures not abandoned	4	2.5
6	Inability to master the process after release by the development group	1	10.5
7	Lack of personnel in research and development	24	23
8	Failures in management; insufficient interest on the part of managers	5	4.5
9	Long time required by managers for coordination	7	1
10	Differences of opinion between managers and experts	3	14
11	Failures in preparation for production	25	25
12	Delay in construction activities	17	19.5
13	High costs; planned economy not reached	8	7
14	Insufficient technological and qualitative level	11	4.5
15	Conservative and obsolete views	18	16
16	Inexact and changing objectives	13	22
17	Delay in recognition of problems; failures in communication	21	19.5
18	Changing demand	26	24
19	State orders limiting the project	22	21
20	Insufficient transfer of know-how between branches	23	26
21	Economizing measures	15	12.5
22	Unfavorable price relations	16	8
23	Insufficient special knowledge	6	17
24	Uncoordinated development among several branches	14	12.5
25	Better solutions from competitors	9	10.5
26	New solutions replacing the initial project	20	15

- Faster recognition of problems and improvement in communication (17)
- Better adaptation to new state orders and laws (19)
- Positive changes in views and approaches (15)
- Reduction in stress caused by other production tasks (3)

3.2.3 An Approach to Finding the Main Determinants of Innovation

Our discussions with managers confirmed that the ability to master the innovation process is a complex phenomenon. Some specialists stress the importance of creative or innovative potential, but if this potential is not used appropriately, the results will be inadequate. A second major factor is thus the firm's long-term strategic orientation. Yet even with considerable potential for innovation and an appropriate strategy, the process might be arrested by stress resulting from other production tasks. The capacity for mastering ongoing processes is therefore a third factor. As the innovation process is complex, touching the entire network of supplier and buyer relations, a fourth factor involves cooperation and coordination. These four determinants are related to at least some extent to the main stages of the innovation process and thus led us to the analytical scheme shown in Table 21.

We adjusted the 26 variables to the four determinants (I, S, O, C) in research and development, production, marketing, and in management at all stages by our assumptions of their dependencies. To prove this we used multivariate factor analysis, which enables us to identify the main factors among many variables by investigating their latent intercorrelation. In this case we used as a criterion the so-called factor loading of a variable at

TABLE 21 Determinants of innovation, as measured by variables, in research and development, production, marketing, and management.

Determinant	Variables: In research and development	In production	In marketing	In management (all stages)
Innovative potential (*I*)	2 11	2	14	6 15
	5 14	6		8 23
	7 26	13		10
Strategic orientation (*S*)	1 17	22	18	9 16
	7			10 17
	14			15
Capacity for mastering ongoing processes (*O*)	3	3	18	8
	7	13		9
		21		10
Cooperation and coordination (*C*)	1	1	20	1 17
	4	4	25	9 19
	24	20		10

a level of at least ±0.40. We were able to identify 7 factors in the case of variables inhibiting innovation (see Table 22) and 7 factors in the case of variables promoting innovation (see Table 23). Adjusting these factors to determinants and stages of the innovation process produces the following results:

A. Variables inhibiting innovation

Factor	*Determinant*	*Stage*
1	Innovative potential (*I*)	Research and development
2	Strategic orientation (*S*)	All (management)
3	Cooperation and coordination (*C*)	Research and development
4	Economic mechanism (*E*)	All (management)
5	Know-how (*K*)	All (management)
6	Cost	All (management)
7	–	All (management)

B. Variables promoting innovation

Factor	*Determinant*	*Stage*
1	Strategic orientation (*S*)	Research and development
2	Cooperation and coordination (*C*)	Research and development
3	Strategic orientation (*S*)	All (management)
4	Cooperation and coordination (*C*)	All (management)
5	Capacity for mastering ongoing processes (*O*)	All (management)
6	Innovative potential (*I*)	Production
7	–	All (management)

While innovative potential, strategic orientation, and cooperation and coordination are the main determinants connected to variables inhibiting innovative activities, innovative potential does not play such an important role on the side of variables promoting innovative

TABLE 22 Variables inhibiting innovation and their factor configurations.

No.	Variable	Loading factor	No.	Variable	Loading factor
Factor 1			*Factor 4*		
11	Failures in preparation for production	0.81	22	Unfavorable price relations	0.75
7	Lack of personnel in research and development	0.69	3	Stress caused by other production tasks	0.74
15	Conservative and obsolete views	0.63	19	State orders limiting the project	0.46
25	Better solutions from competitors	0.62			
19	State orders limiting the project	0.41	*Factor 5*		
			6	Inability to master the process after release by the development group	0.72
Factor 2					
18	Changing demand	0.74	2	Technical difficulties	0.60
16	Inexact and changing objectives	0.70	23	Insufficient special knowledge	0.42
1	Insufficient supply from the supplier industry	0.66			
17	Delay in recognition of problems; failures in communication	0.55	*Factor 6*		
			13	High costs; planned economy not reached	0.62
12	Delay in construction activities	0.52	21	Economizing measures	0.42
			19	State orders limiting the project	0.41
Factor 3					
24	Uncoordinated development among several branches	0.66	*Factor 7*		
21	Economizing measures	0.64	8	Failures in management; insufficient interest on the part of managers	0.67
4	Insufficient supply of machines and means of rationalization	0.61			
			10	Differences of opinion between managers and experts	0.58
26	New solutions replacing the initial project	0.55	5	Developmental failures not abandoned	0.54
12	Delay in construction activities	0.41			

TABLE 23 Variables promoting innovation and their factor configurations.

No.	Variable	Loading factor	No.	Variable	Loading factor
Factor 1			*Factor 4*		
17	Delay in recognition of problems; failures in communication	0.87	4	Insufficient supply of machines and means of rationalization	0.88
20	Insufficient transfer of know-how between branches	0.84	21	Economizing measures	0.62
12	Delay in construction activities	0.64	1	Insufficient supply from the supplier industry	0.41
11	Failures in preparation for production	0.59	13	High costs; planned economy not reached	0.41
15	Conservative and obsolete views	0.42			
1	Insufficient supply from the supplier industry	0.40	*Factor 5*		
			3	Stress caused by other production tasks	0.63
Factor 2			10	Differences of opinion between managers and experts	0.50
25	Better solutions from competitors	0.71	14	Insufficient technological and qualitative level	0.45
24	Uncoordinated development among several branches	0.67	19	State orders limiting the project	0.43
26	New solutions replacing the initial project	0.67	8	Failures in management; insufficient interest on the part of managers	0.42
22	Unfavorable price relations	0.66			
15	Conservative and obsolete views	0.65			
14	Insufficient technological and qualitative level	0.57	*Factor 6*		
			6	Inability to master the process after release by the development group	0.73
Factor 3					
23	Insufficient special knowledge	0.74	4	Insufficient supply of machines and means of rationalization	0.58
16	Inexact and changing objectives	0.68			
10	Differences of opinion between managers and experts	0.66	13	High costs; planned economy not reached	0.50
7	Lack of personnel in research and development	0.63			
18	Changing demand	0.50	*Factor 7*		
14	Insufficient technological and qualitative level	0.45	9	Long time required by managers for coordination	0.68
			8	Failures in management; insufficient interest on the part of managers	0.44

activities. We identified three other important determinants, the economic mechanism (including price relations, planning mechanisms, and other incentives), know-how, and cost. Our improved scheme for factor analysis is shown in Table 24, which illustrates the complexity of innovation management.

3.2.4 Using Factor Profiles in Comparing Enterprises

While the number of innovations analyzed was too small to allow us to draw further conclusions, it became clear to us that the 32 firms we investigated did not sufficiently develop innovative potential. The influence of both factors inhibiting innovation and factors promoting innovation in a given firm can be described by a profile. We also discovered that the objective factor configuration is far more unified than is the specific behavior of firms. This finding suggests that we should pay more attention to the objective factor configuration of the innovation process according to industry, to the national economy, and to basic innovations and improvement-related innovations. On the other hand, we should analyze the individual behavior of firms and compare our results with the objective factor configuration on the level of the industry or society; this could provide us with information, not only about the firm's management, but also about national policy for innovation.

The consequences of an inadequate policy for innovation in an industrial firm are not always immediately apparent. It may also take a long time to develop and to use creative potential. Managers should give the greatest attention to the human factor and to the appropriate combination of important factors. We propose investigating this problem by a specific profile showing the strength of factors inhibiting innovation and of a firm's own activities in promoting innovation during the innovation process. Figure 13 shows such a profile for the sampled 32 firms in sectors of the consumer goods industry.

We note the greatest differences between the strength of factors inhibiting innovative activities and the strength of the firm's own capabilities in the following determinants and stages:

- Cooperation and coordination: research and development
- Innovative potential: production
- Know-how: production
- Capacity for mastering ongoing processes: marketing

Therefore, a long-term development program for a given industry should include measures for improving organization in research and development and for increasing the qualification level in production. Current organizational changes in industry in the GDR have the explicit goal of mastering the complexity of the innovation process and enabling firms to implement their new products and processes without bureaucratic delays. In this process, exchange of experience between enterprises plays an important role.

Comparison of enterprises is an effective tool for recognizing both bottlenecks and opportunities. For example, Fig. 14, which compares a single firm's profile with the average of the investigated sample, shows that the firm under consideration might have experience in marketing that would be useful for other enterprises. Comparison of enterprises was formerly oriented primarily toward technical and economic indicators. Comparison of determinants of the innovation process, innovative potential, and know-how could be a useful addition to these traditional tools of management.

Profiles enable us to trace major gaps and bottlenecks and to discover possible directions for further investigation of obstacles and factors blocking innovative activities, thus

TABLE 24 Determinants and factors at various stages of the innovation process, as measured by 26 variables.

Stage	Determinant and factor													
	1 Innovative potential (*I*)		2 Strategic orientation (*S*)		3 Cooperation and coordination (*C*)		4 Capacity for mastering ongoing processes (*O*)		5 Economic mechanism (*E*)		6 Know-how (*K*)		All	
	p	*q*	*p*	*q*	*p*	*q*	*p*	*q*	*p*	*q*	*p*	*q*	*p*	*q*
Research and development	*11*		7	*17*	*24*	*25*	3				5			
	7		14	*20*	*21*	*24*	7				12			
	15		17	12	*4*	*26*					20			
	25			11	26	22					23			
	24			*15*	12	15								
	19			1	(3)	14								
	(1)			(1)		(2)								
Production	6	*6*	16				3				6			
	13	*14*	20				13				20			
	23	13	24				21				23			
		(6)												
Marketing	14		18				18				14			
							23				18			
All (management)	8		*18*	*23*	*4*	*4*	*13*	*3*	*22*		*6*		*8*	*9*
	10		*16*	*16*	*21*	*21*	21	*10*	*3*		*2*		*10*	*8*
	15		1	10	*1*	*1*	19	14	*19*		*23*		5	(7)
			17	*7*	13	13	(6)	19	(4)		*20*		(7)	
			12	*18*	(4)	(4)		*8*			(5)			
			(2)	14				(5)						
				(3)										

NOTES:
p = variables inhibiting innovative activities.
q = variables promoting innovative activities.
Numbers in parentheses are factor numbers.
Figures in factor fields are variables from multivariate analysis.
Numbers in italics show other appropriate variables.
The cost factor is omitted because it is of little importance.

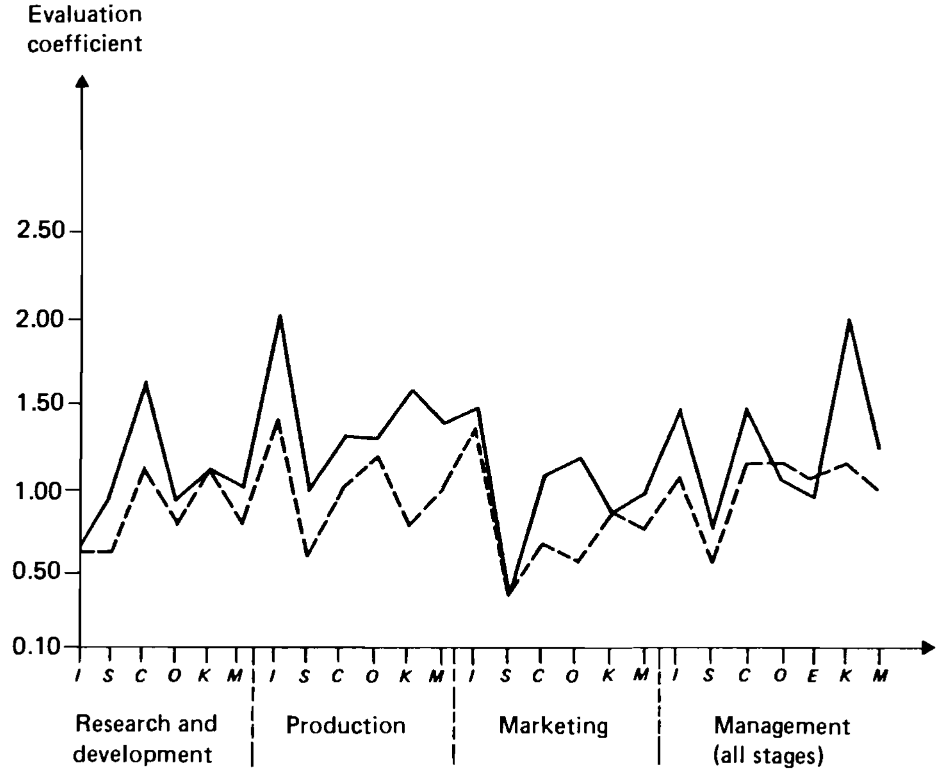


FIGURE 13 Profile of the strength of factors inhibiting innovative activities (——) and factors promoting innovative activities (– – –) in 32 firms (average), where *I* is innovative potential, *S* is strategic orientation, *C* is cooperation and coordination, *O* is capacity for mastering ongoing processes, *E* is economic mechanism, *K* is know-how, and *M* is mean value.

providing an instrument for management at the corporate level. Under a planned economy, exchange of experience and competition between teams of workers in outbidding planned figures play an important role. A firm's profile further explains the quantitative indicators of efficiency. On the other hand, we can assume that profiles show significant differences among branches of industry and among stages of the efficiency cycle. Progression through take-off, rapid growth, maturation, saturation, and decline is connected with structural changes, which should be planned at upper levels of management.

3.3 Innovation and the Efficiency Cycle

Our investigation of the roles of basic and improvement-related innovations and our analysis of the life cycles of industrial organizations can help us to understand better why the innovation process is not continuous as we might first assume; rather, it is interrupted by the effects of stimulation or its absence. The relationship between basic and improvement-related innovations drives the process of technological and economic development. This relationship is at the core of the special circumstances surrounding the birth, growth, and decline of each successive new branch of industry. Simple demand pull models or technology push models are therefore inadequate explanations of the innovation process – in

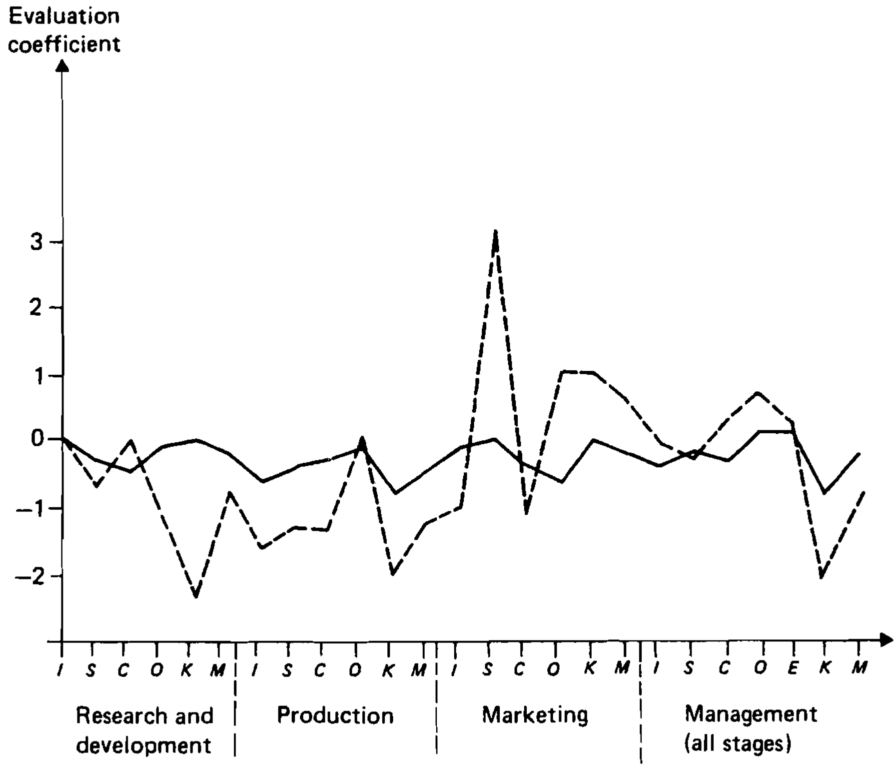


FIGURE 14 Difference between strength of factors inhibiting and strength of factors promoting innovative activities for average of 32 firms (——) and firm 27 (- - -), where I is innovative potential, S is strategic orientation, C is cooperation and coordination, O is capacity for mastering ongoing processes, E is economic mechanism, K is know-how, and M is mean value.

specific branches of manufacturing industry or in the economy as a whole. The interaction between science, technology, and the economy varies in its nature and intensity over time and among various industries.

We cannot say that inventions are always the simple result of demand pull. Need and demand are the main driving factors in the diffusion process. When we look at the innovation process in retrospect, we find that inventions are all caused by an existing need, but the more important ones came from a rather probabilistic cognitive process that led to the achievement of goals that had not previously been realized. Penicillin, saccharin, and synthetic rubber are examples. At the end of the invention process, needs that were not the original targets of research and development were satisfied. Often demand pull is the main reason that incremental innovations use the efficiency potential of basic innovation. But fundamental inventions are less (or not as directly) connected with market demand or concrete needs. Basic innovations create new fields for production and efficiency through, say, a series of new scientific discoveries and technological advances. The connection between these advances and the developing needs of society is often realized slowly. The role of basic innovations in the development of efficiency is demonstrated through Fig. 15.

We turn now to the impact of basic and improvement-related innovations on the economy. Efficiency in general is

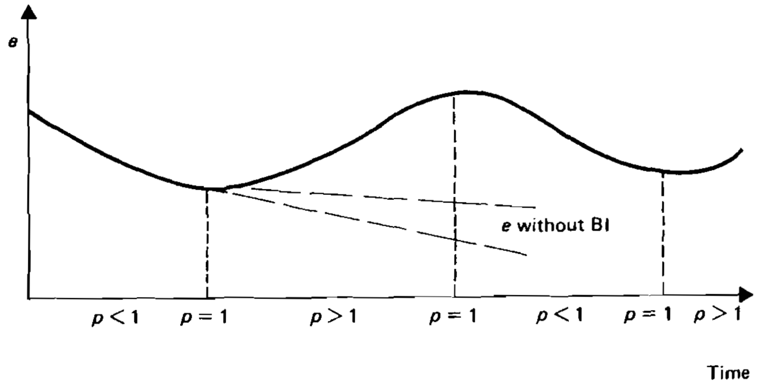


FIGURE 15 Role of basic innovations in the development of efficiency, where e is relative efficiency, BI is basic innovation, and p is given by eqs. (28) and (29).

$$e_0 = E_0/C_0 \tag{22}$$

where

E_0 is the sum of benefits or revenues at time $t = 0$
C_0 is the sum of costs or expenditures at time $t = 0$

At $t = 1$ we find

$$e_1 = \frac{E_0 + \Delta E}{C_0 + \Delta C} = e_0 \frac{1 + \Delta E/E_0}{1 + \Delta C/C_0} \tag{23}$$

$$\Delta E = E_1 - E_0 \tag{24}$$

$$\Delta C = C_1 - C_0 \tag{25}$$

The increase of E can be divided into

$$\Delta E = \Delta E_N + \Delta E_A \tag{26}$$

where

ΔE_N is the increase in benefits or revenues from new processes and products
ΔE_A is the increase in benefits or revenues from old processes and products

At the same time, for costs

$$\Delta C = \Delta C_N + \Delta C_A \tag{27}$$

Therefore

$$e_1 = e_0 \frac{1 + (\Delta E_N + \Delta E_A)/E_0}{1 + (\Delta C_N + \Delta C_A)/C_0} \tag{28}$$

$$e_1 = e_0 p \tag{29}$$

A pure improvement policy gives us $\Delta E_N = 0$ and $\Delta C_N = 0$. However, initially we have high benefits ΔE_A in connection with moderate expenditures ΔC_A, with $p > 1$. Later, we have diminishing returns and thus $p < 1$ and a decrease in efficiency.

A pure or dominant policy of improvement leads to a situation described by many authors as a "productivity dilemma," where primary attention is given to short-term gains and new basic innovations do not occur or are delayed. The inertia of the given technological system becomes a major barrier for further economic progress. Therefore, efficiency e declines because of a lack of gains from substantial improvement-related innovations, which may be explained by the inevitable increase in costs for resources, environment, and infrastructure.

This situation is critical for further economic development. If we are unable to stimulate inventions that can open new directions and fields of economic activity and thus improve efficiency, the result will be predictable: a decline in the ability to meet national and societal needs, to overcome shortages of resources, to avoid unemployment, and to promote the conditions necessary for business activity, especially investment activity. In the case of $p < 1$, the innovation process has "run dry" owing to the effects of innovations, which have no positive influence on efficiency, or to improvement-related innovations that cannot compensate for increasing costs. The result is stagnation or resource crises with grave social and political consequences – crises very different from the usual, seven- to ten-year ups and downs in the business cycle of capital reproduction.

4 CONCLUSIONS

In this report we have dealt with the concept of relative efficiency. Clearly, innovative efficiency can be measured only by measuring the efficiency of the innovating system, which can be better understood by comparison over time with the efficiency of the next higher system. When there are considerable changes in the efficiency of a national economy as a whole, there are also essential changes in the relative efficiency of a given set of innovations. Changes in the prices of resources have a direct and indirect impact on the efficiency of innovations. Price increases lower the absolute efficiency of innovating systems, but may, on the other hand, improve relative efficiency in some cases.

A dynamic view of efficiency is important for innovation policy. In both market and planned economies, it is necessary to see, not only the bottlenecks and shortcomings, but also the prospects and opportunities offered by a given stage in the efficiency cycle. The efficiency cycle is a challenge for management. Managers should be able and ready to change their approaches according to the requirements of various stages and to master growing complexity of innovation management. We found that the combination of factors influencing innovative efficiency changes considerably and over the entire cycle shifts from the innovating system to the environment. The better the innovating system can master outside problems, the more likely that the innovation will succeed. Here we find also the explanation for the striking differences in the importance of factors influencing innovations. Sometimes the principal factors are entrepreneurship and the role of managers; sometimes, the understanding of user needs and marketing; and sometimes, managerial techniques and strength.

Our analysis of different sets of factors revealed that innovative potential, strategic orientation, cooperation and coordination, and know-how are the main determinants for the success of innovations from the standpoint of an industrial organization. These factors should be developed for each stage in the efficiency cycle. Our investigation could not determine the concrete interdependence of these determinants in various stages; this will be a goal for further research. It is obvious, however, that the interface of determinants must be described in terms of policy making and active response to the needs of the given stage in the cycle.

We also think that such traditional means of economic analysis as productivity analysis should be revised from the standpoint of the efficiency cycle and the influence of innovations. Such an analysis must have a more efficient forecasting power and operational value, which can be reached if the analysis answers three questions:

- What change in efficiency can we expect in the future?
- How can we master the requirements of the next stage in the efficiency cycle of the given system?
- How can we combine innovating and noninnovating subsystems to ensure stable growth of efficiency in the industry as a whole?

ACKNOWLEDGMENTS

We would like to extend our thanks for fruitful discussions and useful hints to Professor Daniel Roman, the George Washington University, Washington, DC; Professor Walter Goldberg, International Institute of Management, Berlin (West); Professor Rolfe Tomlinson, School of Industrial and Business Studies, University of Warwick, UK; Professor Walter Zegveld, Netherlands Organization for Applied Scientific Research (TNO), Delft, the Netherlands; and Professor Nikolai Lapin, Academy of Sciences, USSR.

REFERENCES

Abernathy, W.J. (1978) The Productivity Dilemma: Roadblock to Innovation in the Automobile Industry. Baltimore and London: Johns Hopkins University Press.

Abernathy, W., and J. Utterback (1978) Patterns of Industrial Innovation. Technology Review 80(7): 40–47.

Eisendraht, E. (1964) Brutto- und Nettoprinzip in der westlichen und östlichen Wirtschaftsforschung. In Vierteljahreshefte für Wirtschaftsforschung. Berlin (West): Duncker und Humblot.

Energy Systems Program Group of the International Institute for Applied Systems Analysis (1981) Energy in a Finite World: A Global Systems Analysis. W. Häfele, Program Leader. Cambridge, Massachusetts: Ballinger.

Gold, B., ed. (1975) Technological Change: Economics, Management and Environment. Oxford: Pergamon Press.

Goldberg, W. (1980) Stagnation and Innovation: Relevant Policy Questions. DP-80-12. Berlin (West): International Institute of Management.

Haustein, H.-D. (1974) Die Proportionalität der technischen Basis. Berlin (East): Akademie Verlag. In German.

Haustein, H.-D. (1976) Messung der volkswirtschaftlichen Intensivierung. Berlin (East): Akademie Verlag. In German.
Haustein, H.-D. (1979) Innovations in the Lighting Industry. WP-80-12. Laxenburg, Austria: International Institute for Applied Systems Analysis.
Henriot, I. (1955) Geschichte des Schiffbaus. Leipzig-Jena: Urania Verlag. In German.
Knight, K.E. (1967) A Descriptive Model of the Intra-Firm Innovation Process. Journal of Business 40, October:478–496.
Krengel, R., E. Baumgart, A. Boness, R. Pischner, and K. Droege (1973, 1975, 1978) Produktionsvolumen und -potential: Produktionsfaktoren der Industrie im Gebiet der BRD. Statistische Kennziffern. Berlin (West): Deutsches Institut für Wirtschaftsforschung.
Maier, H. (1979) Some Conclusions for International Technology Assessment from the IIASA Energy Study. Presented at the Twenty-fourth International Meeting of the Institute of Management Sciences, Honolulu, Hawaii, 18–21 July, 1979.
Mensch, G. (1976) Gemischtwirtschaftliche Innovationspraxis. Göttingen: Otto Schwartz. In German.
Myers, S., and D.G. Marquis (1969) Successful Industrial Innovations: A Study of Factors Underlying Innovation in Selected Firms. NSF 69-17. Washington, DC: US National Science Foundation.
Nelson, R.R., and S.G. Winter (1977) In search of useful theory of innovation. Research Policy 6(1): 36–76.
Nick, H. (1974) Die Wechselwirkung revolutionärer und evolutionärer Veränderungen. In E. Gatovskij, H. Maier, F. Kutta, Wissenschaft, Technik und Produktion. Wissenschaft, Ökonomie, Fortschritt. Berlin (East): Akademie Verlag. In German.
OECD Economic Outlook (1979)25: July.
Ray, G.F. (1979) Der Innovationsprozess in Westeuropäischen Industrieländern. Innovation in der Energiewirtschaft. Band 3. Berlin (West): Duncker und Humblot. In German.
Schumpeter, J.A. (1952) Theorie der wirtschaftlichen Entwicklung. Berlin (West): Duncker und Humblot. In German.
Science Policy Research Unit of the University of Sussex (1972) Success and Failure in Industrial Innovation: Report on the Project SAPPHO. London: Centre for the Study of Industrial Innovation.
Statistical Yearbook of the USSR (1977) Moscow: Statistika.
Uhlmann, L. (1978) Der Innovationsprozess in Westeuropäischen Industrieländern. Band 2. Der Ablauf industrieller Innovationsprozesse. Berlin (West): Duncker und Humblot. In German.
US National Science Board (1977) Science Indicators 1976: Report of the National Science Board. 1977. NSB-77-1. Washington, DC: US Government Printing Office.
US Energy Research and Development Administration (1976) A National Plan for Energy Research, Development and Demonstration: Creating Energy Choices for the Future. ERDA-48. Washington, DC: US Government Printing Office.
Zvorykin, A.A., N.I. Osmova, V.I. Chernyshev, and S.V. Shukhardin (1962) Istorija tekhniki. Moscow: Gospolitizdat. In Russian.

APPENDIX A A Method for Calculating the Time Needed to Equalize Productivity Levels in Two Countries

Given

- The productivity level Y_t of countries A and B at time $t = 1$
- The average productivity growth rate λ_m for time period $t = 1, 2, \ldots, m$ in both countries
- The expected future productivity growth rate λ_r for time period $t = m, m + 1, \ldots, m + r$

We can assume

$$Y^A = Y_1^A(1 + \lambda_m^A)^{m-1}(1 + \lambda_r^A)^r \tag{A1}$$

$$Y^B = Y_1^B(1 + \lambda_m^B)^{m-1}(1 + \lambda_r^B)^r \tag{A2}$$

The growth rate is

$$\lambda_m = (Y_m/Y_1)^{1/m-1} - 1$$

From eq. (A1) = eq. (A2) we arrive at

$$r = \frac{\ln(Y_1^A/Y_1^B) + (m-1)\ln[(1+\lambda_m^A)/(1+\lambda_m^B)]}{\ln[(1+\lambda_r^B)/(1+\lambda_r^A)]} \tag{A3}$$

This equation is meaningful if

$$Y_1^A > Y_1^B \tag{A4}$$

$$\lambda_r^B > \lambda_r^A \tag{A5}$$

$$\Delta\lambda = \lambda_r - \lambda_m \tag{A6}$$

and

$$\Delta\lambda^B > \Delta\lambda^A \tag{A7}$$

$$r = \frac{\ln(Y_1^A/Y_1^B) + (m-1)\ln[(1+\lambda_m^A)/(1+\lambda_m^B)]}{\ln[(1+\lambda_m^B+\Delta\lambda^B)/(1+\lambda_m^A+\Delta\lambda^A)]} \tag{A8}$$

Substituting

$$\ln(Y_1^A/Y_1^B) = C \tag{A9}$$

$$1 + \lambda_m = i_m \tag{A10}$$

We can write eq. (A7) as follows:

$$r = \frac{C + (m-1)\ln(i_m^A/i_m^B)}{\ln[(i_m^B+\Delta\lambda^B)/(i_m^A+\Delta\lambda^A)]} \tag{A11}$$

In practical cases, C, m, i_m^A, and i_m^B are given. For $\Delta\lambda^B$ and $\Delta\lambda^A$ we can assume a first estimation over time period r^*. By calculating eq. (A11), we arrive at a first approximation of r_1. Then we must compare r_1 and r^*. If $r_1 \leqslant r^*$, there is obviously no problem. But if $r_1 > r^*$, we should decide whether there is any reason for improving the $\Delta\lambda$ assumed. We then arrive at a second approximation of r.

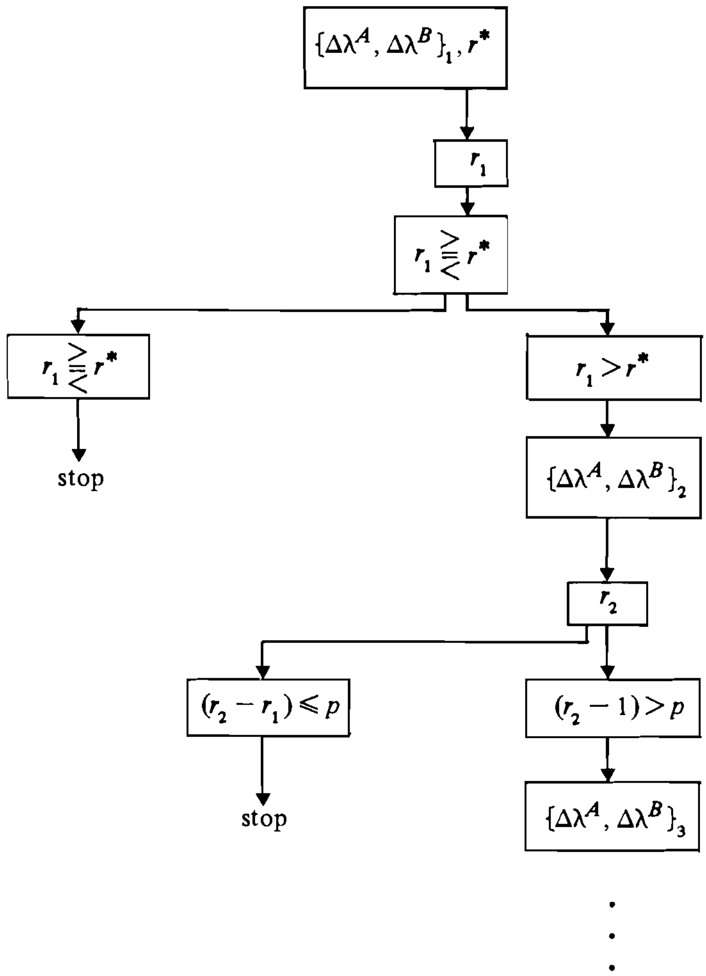
Iteration continues until

$$r_n - r_{n-1} \leqslant p$$

We assume that

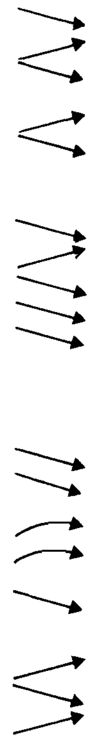
$p = 1$ in forecasts of 1 to 5 years
$p = 2$ in forecasts of 6 to 10 years
$p = 5$ in forecasts of 11 to 20 years
$p = 10$ in forecasts of more than 20 years

The iteration process is as follows:



APPENDIX B The Efficacy of Factors Influencing Innovative Activities

Level and factor	Hypothesis	Type of change[a]	Schema of efficacy (RDAE)[b]
I. Level of innovator			
a. Input/output			
a1. Input-related factors			
Labor	The need for skilled labor decreases.	A2	
Capital equipment	The technology becomes increasingly capital-intensive (mechanized).	A1	
Raw materials, components, services, facilities	Things required become generally available.	A2	
Natural assets	Scaling up the technology endangers the natural environment.	A1	
Knowledge from other innovation processes	Knowledge required becomes increasingly available.	A2	
a2. Output-related factors			
Technological risk of failure	The technology becomes increasingly controllable.	A2	
Complexity	Technological features become increasingly comprehensive and interdependent.	A3	
Communicability, clearness	Characteristics of the technology become increasingly known and understandable.	A3	
Scale of one unit	Divisibility is increasingly promoted.	A3	
Usability as an element of a more comprehensive technological system	Applicability is increasingly ensured.	A3	
b. Interaction among innovators			
b1. Interplay of functional roles			
Initiator	The impulse to develop the technology is decisive.	A2	
Expert	Competence required to find a solution to emerging problems becomes increasingly ubiquitous.	A2	
Helper	Impetus is required primarily during the transition from exploration to investment.	A1	
Stabilizer	Implementing a solution and integrating it into the existing system are most difficult during application.	A2	
Staffing	Matching roles and persons can be achieved by trial and error.	A2	
b2. Characteristics of innovators			
Personal interest	Boredom increases with routinization.	B3	
Experience	Practice and skill required become increasingly available.	A2	
Creativity	Boredom and exploitation of technological opportunities reduce the probability that new ideas will emerge.	B3	



II.	Level of organization			
c.	Resources			
	Labor	Dependence on the availability of manpower decreases.	A2	↘
	Capital equipment	The technology becomes increasingly capital-intensive (mechanized).	A1	↗
	Financial funds	Funds needed increase because of scaling up.	A1	↗
	Natural assets	Scaling up the technology endangers the natural environment.	A1	↗
	Raw materials, components, services, facilities	Things required become generally available.	A2	↘
	Infrastructure of facilities	Facilities are generally improved and adjusted to requirements.	A2	↘
	Access to knowledge	Knowledge required becomes increasingly available.	A2	↘
d.	Organizational dimensions			
d1.	Relationships with the environment			
	Branch of industry	The influence of the branch alone cannot be estimated.	–	⇝
	Position in market	The influence of a strong or a weak position alone cannot be estimated.	–	⇝
	Response to client's needs	Needs and applicability of the technology match more and more.	A2	↘
	Response to technological trajectories	Future technological features become increasingly clear.	A2	↘
	Response to economic and social trajectories	Future economic and social features become increasingly clear.	A2	↘
d2.	Internal dimensions			
	Size of the organization	The influence of size alone cannot be estimated.	–	⇝
	Ownership, influence of mother companies	The influence of ownership alone cannot be estimated.	–	⇝
	System of goals	The technology is increasingly included in the organization's system of goals.	A2	↘
	Technological and innovative philosophy (innovativeness)	The technology's compatibility with the organization's philosophy is examined during the early stages of the innovation process.	A2	↘
	Decision-making principles	The influence of the decision-making habit alone cannot be estimated.	–	⇝
	Principles of budgeting	Problems related to including the technology in the budget become increasingly easily solved.	A2	↘
	Principles of planning	Increasing clearness makes planning easier.	A2	↘
	System of incentives	Labor input becomes decreasingly important.	A2	↘
	System of sanctions	Imposing penalties in the case of inability to promote innovation becomes more important as work progresses.	A1	↗
	Job allocation	Combining manpower and activities becomes increasingly easy.	A2	↘
	Principles of training and professional development	Problems occur primarily during the early stages of the innovation process.	A1	↘

APPENDIX B *continued.*

Level and factor	Hypothesis	Type of change[a]	Schema of efficacy (RDAE)[b]
Internal social climate	Problems related to the internal social climate that may arise when the technology becomes important may be settled by its acceptance.	B3	
Formal organizational principles	Implementing the technology may cause problems in the existing organizational structure.	A1	
e. Organizational measures			
e1. Planning measures			
Objectives in pursuing the technology	The influence of the objectives alone cannot be estimated.	–	
Selection of projects	The more the technology is understood, the more easily decisions may be made.	A3	
Evaluation and forecasting of costs and benefits	The more experience is gained, the less difficult the problem is.	A2	
Determination of interaction among innovators	Assigning roles to persons is more difficult during the early stages of the innovation process.	A2	
Target setting for innovators	Regulating time and resources becomes easier as the technology becomes clearer.	A3	
e2. Control measures			
Supervision of innovative activities	The need for and difficulties of supervision are greatest in "open" stages, where there are several ways to proceed.	A2	
Decision on solutions proposed by innovators	A decision's influence depends on its determinants.	–	
Utilization of knowledge acquired	The clearer the technology becomes, the easier it is to decide how to use and transfer knowledge.	A2	
III. Level of environment			
f. Resources			
Labor	The need for skilled labor decreases.	A2	
Capital equipment	The technology becomes increasingly capital-intensive (mechanized).	A1	
Financial funds	Funds needed increase due to scaling up.	A1	
Natural assets	Scaling up the technology endangers the natural environment.	A1	
Raw materials, components, services, facilities	Things required become generally available.	A2	
Infrastructure of facilities	Facilities that the environment must provide become increasingly important as the technology is scaled up.	A1	
Funds of and access to knowledge	Knowledge required becomes increasingly available.	A2	

g.	Environmental dimensions			
g1.	Economic sector			
	Size of the economy	The influence of size alone cannot be estimated.	–	⇝
	Degree of division of labor	The influence of increasing specialization alone cannot be estimated.	–	⇝
	System of competition	Competition becomes increasingly relevant as the technology is scaled up.	A1	↗
	International economic engagement	The influence of international trade alone cannot be estimated.	–	⇝
g2.	Political sector			
	National goals	Consistency with the system of national goals becomes increasingly relevant as the technology requires more and more resources.	A1	↗
	Allocation of political power	The influence of the allocation of competence cannot be estimated.	–	⇝
	System of facilities for research and development	Influence is especially relevant during research and development.	A1	↘
g3.	Social sector			
	System of social values	Social values become increasingly relevant as the technology is exposed to the public.	A1	↗
	Social attitude toward new technologies	Social acceptance becomes increasingly relevant as the technology is exposed to the public.	A1	↗
h.	Environmental measures			
h1.	Economic sector			
	Development of demand, economic prospects	Demand must first be created; it is then matched and is increasingly exhausted.	B1	⤴
	Competing technologies	Advantages of the technology against those of other emerging technologies are increasingly lost.	B2	↗
	Cooperation with suppliers	Dependence on suppliers' know-how decreases.	A2	↗
	Cooperation with buyers	The need to inform and help buyers is most important during the early stages of application.	A2	⤵
	Agreements with competing firms	The difficulties and effects of making agreements with competing firms cannot be estimated.	–	⇝
	Wages	Wages have their greatest influence during the early stages of application.	A1	⤵
	Prices of capital equipment	The more capital-intensive the technology becomes, the more important the costs of capital usage are.	A1	↗
	Prices of raw materials, components, services, facilities	Dependence on the costs of goods input increases with the emergence of competing firms and technologies.	B2	↗
	Environmental effects	Scaling up the technology leads increasingly to problems with the natural environment.	A1	↗
	Revenues achievable from the technology	After passing the break-even point, profits decrease with augmenting exploitation.	B1	⤴

APPENDIX B *continued.*

Level and factor	Hypothesis	Type of change[a]	Schema of efficacy (RDAE)[b]
h2. Political sector			
Regulations	Regulations imposed by authorities are most relevant during the early stages of application.	A2	
Taxes	Taxation becomes increasingly relevant as the technology is scaled up.	A1	
Government grants	Governments subsidize the "middle" phase of the innovation.	A1	
Government incentives	The influence of government incentives alone cannot be estimated.	–	
h3. Social sector			
Public familiarity with the technology	The technology becomes increasingly known to the public.	A1	
Attitude toward the technology	Social acceptance of the technology becomes increasingly relevant.	A1	
Professional awards, prizes, and image	Lack of recognition is most restrictive in the early stages.	A1	

[a]See Sections 2.2.2 and 2.3.2 for explanation.
[b]See Sections 2.3.1 and 2.3.2 for explanation.

CHANGES IN COMPARATIVE ADVANTAGES AND PATHS OF STRUCTURAL ADJUSTMENT AND GROWTH IN SWEDEN, 1975–2000

Lars Bergman and Lennart Ohlsson
Stockholm School of Economics, Sweden

SUMMARY

The purpose of this study is to identify possible future development paths for the Swedish economy in a context where world market conditions, domestic factor accumulation, and technical change are explicitly taken into account. The main analytical tool used in the study is a general equilibrium model of the Swedish economy. World market prices and trade flows as well as domestic factor accumulation and productivity change are exogenous to the model. The sectoral allocation of capital and labor as well as domestic consumption, foreign trade, and the domestic price system are endogenously determined variables.

The study's projections indicate that Sweden is entering a period of considerably slower economic growth than occurred during the earlier part of the postwar period. Underlying this result is an assumed slowdown of the productivity growth rate. The assumed rates of productivity change do not differ significantly between the sectors. Consequently, reallocation gains can be achieved mainly through a reduction of the intersectoral differences in the marginal productivity of capital, characterizing the initial year of the projection period.

1 BACKGROUND AND AIM OF THE STUDY

The research presented in this report is inspired by the slowdown of economic growth and the emergence of new "problem" industries and regions in Sweden, as in many other industrialized countries, during the 1970s. Only to some extent do these problems seem to be of a short-term, business cycle nature. One of several long-term reasons might be a sustained gradual shift in the pattern of comparative advantages of industrialized countries. There may be many possible reasons behind such a shift. One is that developing countries are becoming increasingly competitive in several markets where industrialized countries previously dominated as suppliers. Other reasons are, for instance, differential growth rates among countries, differential rates of factor accumulation, and differential technical changes among sectors. Changes in the internal functioning of the economy, however, may also have contributed to a bad aggregate performance.

In some, and perhaps most cases, the sources of comparative advantage changes in the long run tend to bring about increased productivity of the world economy as a whole. In the short run, however, changes in comparative advantages induce structural adjustment in national economies. If this adjustment is significant, the problems that arise might be, or at least might seem to be, larger than the potential long-term benefits of a complete adjustment to the new pattern of comparative advantages. Moreover, the individual country does not necessarily gain from the comparative advantage changes even in the long run.

The experiences of the Swedish economy in the 1970s are often interpreted as a partial or temporary loss in the ability to adjust rapidly to changing external conditions. Whether this is true or not, Swedish economic policy in the past few years has been largely redirected to ensure that the reallocation of capital and labor from stagnating to expanding industries does not lead to increased unemployment at national, regional, and sectoral levels. (See Ohlsson 1980a, for an analysis of Swedish industrial, labor market, and regional policies with respect to their possible resource allocation effects.)

Policies with such far-reaching aims easily lead to inefficient use of the economy's resources. If they are carried out on a large scale, conflicts are likely to emerge between goals related to economic growth and those related to regional and local employment. One way of reducing the significance of these problems is to create a system of "early warning signals." The rationale of such a system is that if changes in comparative advantages can be foreseen reasonably well, much of the necessary adjustment is taken care of by "normal" market forces and is carried out gradually over an extended time period. Moreover, in such a case there is a better chance that policies for structural change, compatible with various social goals, can be designed and implemented early enough to become efficient and thus reduce demand for protectionism.

Obviously it is not possible to foresee the future. But it is possible to design forecasting methods that are focused on important factors for the development of comparative advantages and that can provide insights into the long-term adjustment behavior of the economy. This is particularly important in economies, like Sweden, that have a large foreign trade dependence but a limited influence on world market conditions.

So far, however, long-term forecasting in Sweden has been focused on capital accumulation, labor supply, and productivity growth. Obviously such factors are very important determinants of economic development, especially if producers face a world market situation that can be characterized as a "seller's market," as was the case in the 1950s and 1960s.

In this paper we nevertheless switch the focus to the development of externally induced comparative advantage changes. This switch is partly motivated by the increasing degree of price competition on world markets, but is also made to find out how external and internal changes in Sweden's comparative advantages interrelate and affect the long-term performance of the economy with regard to a particular policy interest.

Consequently, the purpose of this study is to identify possible future changes in Sweden's comparative advantages and to analyze how these changes might affect the rate and pattern of full employment economic growth, particularly in terms of the sectoral and regional composition of employment. More specifically, we analyze how Sweden's comparative advantages might be affected by specified development paths for world market prices and trade flows and what a complete and smooth adjustment to changing comparative advantages would mean in terms of changes in the sectoral and regional composition of production and employment. In addition, we analyze to what extent alternative scenarios for capital accumulation, labor supply, and productivity growth make significant differences to these dimensions. Apart from highlighting these substantive issues, we develop an approach to the long-term forecasting of comparative advantage changes in a small, open economy.

2 THE MODEL

The model used in the analysis is a computable general equilibrium model of a small, open economy. It belongs to the "family" of such models, which are fully described in Bergman and Pór (forthcoming). Since it is a pure equilibrium model, it does not explicitly incorporate various obstacles to structural change, reflecting the short-run rigidities in capital and labor markets. Thus the main output of the model analysis is a set of conditional estimates of the structural changes of the Swedish economy that would result from a complete adjustment to changes in comparative advantages over a period of 15–25 years.

The model does not have an explicit regional dimension. Thus the regional impact analysis has to be carried out by means of exogenous information concerning the regional distribution of the production units of sectors identified in the model.

In this section the basic structure of the model is briefly described, as are the modifications of the model made for this particular study. For brevity, however, some aspects of the model (for instance the treatment of indirect taxes and tariffs) are simply left out. The growth of the labor force as well as net capital formation for the economy as a whole are exogenous to the model. The same applies to technical change and world market conditions in terms of international prices and production of traded goods in the rest of the world. Thus for a given point in time, world market conditions and the domestic supply of capital and labor are given.

In the model, 23 production sectors and 20 groups of traded goods are identified. In each production sector, capital, labor, fuels, and electricity are substitutable factors of production, whereas the use of non-energy intermediate inputs is proportional to output. The technology exhibits constant returns to scale. The model determines endogenously a sectoral allocation of labor and capital, consistent with equilibrium on all commodity and factor markets at prices equal to marginal (and average) production costs. Accordingly, production, consumption, foreign trade, and price formation are endogenous to the model. By connecting solutions for different points in time, a development path for the economy can be generated.

The model describes an open economy that is "small" in the sense that it faces an elastic supply of imports at parametric prices and cannot influence the export prices of competing countries. In general, however, products with a given classification supplied by domestic producers are treated as imperfect substitutes for products with the same classification supplied by producers in other countries. This approach, which is due to Armington (1969), implies that users of products of a given classification, in the "home country" and elsewhere, actually use a composite of imported and domestically produced goods of that particular classification. The function determining the composition of the composite good, following Armington, is assumed to be homothetic. Moreover, domestic users are all assumed to minimize the unit cost of each type of composite good.

The adoption of this so-called Armington assumption has several implications. One is that there will not be complete specialization in the trade-exposed part of the economy, even though the number of tradable goods exceeds the number of factors of production, and the technology exhibits constant returns to scale. Another is that there will be intra-industry trade. A third implication is that the "home" country will have some influence on its own export prices.

The model describes an economy with $n + 3$ production sectors producing $n + 3$ goods of which n are tradables. There is no joint production, and each good is produced in one sector only. The production sectors are numbered from 0 to $n + 2$, 0 being the electricity sector and 1 the fuels production sector. Since this study is not primarily concerned with energy issues, however, the fuels and electricity sectors are aggregated into one energy sector with index 1. Sector $n + 1$ is the housing sector and $n + 2$ the public sector. There is also a "bookkeeping" sector, $n + 3$, in which different goods are aggregated into one single capital good. Since the number of production sectors is 23 in this particular application, n is set to 21.

Assuming competitive conditions, the prices, P_j, of domestically produced goods are equal to their unit production costs. Thus

$$P_j = \kappa_j(P_1^{\mathrm{D}}, \ldots, P_i^{\mathrm{D}}, \ldots, P_n^{\mathrm{D}}, W_j, R_j, t) \qquad j = 1, 2, \ldots, n+2 \tag{1}$$

where $\kappa_j(\cdot)$ is the unit cost function, and P_i^{D} the price of composite good i, W_j the wage rate in sector j, R_j the user cost of capital in sector j, and t a time index. The heterogeneity of labor is roughly accounted for by an exogenous wage structure, i.e.,

$$W_j = \omega_j W \qquad j = 1, 2, \ldots, n+2 \tag{2}$$

where W is a general wage index and ω_j are constants. The user cost of capital is defined by

$$R_j = P_{n+3}(\delta_j + R) \qquad j = 1, 2, \ldots, n+2 \tag{3}$$

where P_{n+3} is the price of the aggregated capital good, δ_j the rate of depreciation in sector j and R the real rate of interest. The price index of capital goods is defined by

$$P_{n+3} = \sum_{i=1}^{n} P_i^{\mathrm{D}} a_{i,n+3} \quad ; \quad \sum_{i=1}^{n} a_{i,n+3} = 1 \tag{4}$$

As a consequence of the technology assumptions, the unit cost function $\kappa_j(\cdot)$ can be written

$$\kappa_j(\cdot) = \kappa_j^*(P_1^{\mathrm{D}}, W_j, R_j, t) + \sum_{i=2}^{n} P_i^{\mathrm{D}} a_{ij} + Q_j \bar{b}_j \qquad j = 1, 2, \ldots, n+2 \tag{5}$$

where the first part reflects the minimum cost of energy, labor, and capital per unit of output, and the last two parts reflect the cost of non-substitutable inputs per unit of output. Thus the constants a_{ij} represent the input of composite good i per unit of output in sector j, and $\bar{b}_j$ is the corresponding parameter for complementary imports. The world market price, $\bar{Q}_i$, of complementary imports is expressed in the domestic currency unit.

The "net unit cost" function $\kappa_j^*(\cdot)$ is derived from a nested Cobb–Douglas–CES production function, where energy, labor, and capital are variable inputs. Thus there is a constant elasticity of substitution between a composite capital–labor input, defined by a Cobb–Douglas function, and energy. In the original model the aggregated energy good is replaced by a composite fuels–electricity input, defined by a CES function.

The equilibrium prices of the composite goods are given by the unit cost functions of the composites:

$$P_i^{\mathrm{D}} = \phi_i(P_i, P_i^{\mathrm{M}}) \qquad i = 1, 2, \ldots, n \tag{6}$$

where P_i^{M} is the exogenously given world market price of import good i in the domestic currency unit.

Having defined all prices in the model and the unit cost functions $\kappa_j^*(\cdot)$ and $\phi_i(\cdot)$, the derivation of the model is straightforward. Thus there are two types of demand for composite goods: intermediate demand and final demand by the household sector. In addition there is export demand for production sector outputs.

By Shephard's lemma and the assumptions regarding technology, intermediate demand is given by

$$X_{ij} = \begin{cases} \dfrac{\partial \kappa_j^*}{\partial P_i^{\mathrm{D}}} X_j \text{ when } i = 1 \\ \\ a_{ij} X_j \text{ when } i = 2, 3, \ldots, n \end{cases} \qquad j = 1, 2, \ldots, n + 3 \tag{7}$$

Household demand is given by

$$C_i = C_i(P_1^{\mathrm{D}}, \ldots, P_n^{\mathrm{D}}, P_{n+1}, E) \qquad i = 1, 2, \ldots\ n + 1 \tag{8}$$

where E is total household expenditures. In the original model, functions $C_i(\cdot)$ are derived from a utility function such that the resulting demand equations can be represented by a linear expenditure system estimated on the basis of 10 consumer commodity groups and a matrix defining each of the consumer commodity groups as a convex combination of composite goods. Lack of data, however, prevented the use of that version of the model. Instead a system of demand equations with constant expenditure shares for each of the composite goods in household consumption was used. Observe that household demand for energy is derived from the demand for housing services, that is, C_{n+1}.

As a consequence of the Armington assumption, foreign demand for domestically produced goods can be written

$$Z_i = Z_i(P_i, P_i^{\mathrm{W}}, t) \qquad i = 1, 2, \ldots, n \tag{9}$$

where P_i^W is the exogenously given world market price, in the domestic currency unit, of goods with the classification i. In the model it is assumed that the trade-off between goods with different origins is represented by a CES function. Consequently, the function $Z_i(\cdot)$ becomes

$$Z_i = A_i(P_i/P_i^W)^{\epsilon_i} e^{\sigma_i t} \qquad i = 1, 2, \ldots, n \tag{9'}$$

where A_i is a constant, σ_i is the annual rate of change of production of good i in "the rest of the world," and ϵ_i is an elasticity of substitution parameter.

On the basis of Shephard's lemma the equilibrium conditions for the product markets can be written

$$X_i = \frac{\partial \phi_i}{\partial P_i} \left(\sum_{j=1}^{n+3} X_{ij} + C_i \right) + Z_i \qquad i = 1, 2, \ldots, n \tag{10}$$

$$X_i = C_i \qquad i = n+1, n+2 \tag{11}$$

$$X_{n+3} = I + \sum_{j=1}^{n+2} X_j \frac{\partial \kappa_i^*}{\partial R_j} X_j \tag{12}$$

where C_{n+2} is the exogenously given public consumption, and I is the exogenously given net investments.

The demand for competitive imports is given by*

$$M_i = \frac{\partial \phi_i}{\partial P_i^M} \left(\sum_{j=1}^{n+3} X_{ij} + C_i \right) \qquad i = 1, 2, \ldots, n \tag{13}$$

Since $\phi_i(\cdot)$ is derived from a CES function, eqs. (10) and (13) yield the following expression for competitive imports

$$M_i = B_i(P_i/P_i^M)^{\mu_i}(X_i - Z_i) \qquad i = 1, 2, \ldots, n \tag{13'}$$

where B_i is a constant, and μ_i is the elasticity of substitution between imports and domestically produced goods with the classification i. With this formulation the symmetry between the export and import functions becomes obvious. The formulation also shows that here, the small-country assumption implies $X_i - Z_i \approx X_i$ in the rest of the world, i.e., the small country's imports are negligible in relation to production in the rest of the world.

*When solving the model, the functions $\phi_i(\cdot)$ are approximated so that $\partial\phi_i/\partial P_i + \partial\phi_i/\partial P_i^M = 1$. This simplifies some expressions and leads only to minor approximation errors.

Current account equilibrium implies

$$\sum_{i=1}^{n} P_i Z_i = \sum_{i=1}^{n} P_i^{\mathrm{M}} M_i + Q_1 \bar{b}_1 X_1 + D \tag{14}$$

where D is an exogenous variable representing imports to the electricity sector, net transfers, and net interest payments. Observe that complementary imports are used in the energy sector only, the main item being crude oil.

Since capital and labor are inelastically supplied, the equilibrium conditions for the factor markets become

$$K = \sum_{j=1}^{n+2} \frac{\partial \kappa_j^*}{\partial R_j} X_j \tag{15}$$

$$L = \sum_{j=1}^{n+2} \frac{\partial \kappa_j^*}{\partial W_j} X_j \tag{16}$$

where K is capital and L is labor.

After some appropriate substitutions these expressions yield $6n + 10$ equations in the $6n + 10$ unknowns: $X_1, \ldots, X_{n+3}$; $C_1, \ldots, C_{n+1}$; $Z_1, \ldots, Z_n$; $M_1, \ldots, M_n$; $P_1, \ldots, P_{n+3}$; $P_1^D, \ldots, P_n^D$; E; W; and R. Thus the model is determinate. It should be added that the price system is normalized so that the general price level is kept constant over time.

3 SECTORAL CLASSIFICATION AND SCENARIOS

In order to apply the projection model described in the preceding section to the present context, two requirements should be satisfied. The first is that the sectoral breakdown should be consistent with both the theoretical principles underlying the model and the problem focus of the empirical analysis. In the first subsection below the sectoral breakdown used in the study is presented and discussed against the background of this requirement.

The second requirement is that an empirical basis for the definition of exogenous variables and parameters of the model can be established. In order to understand the outcome of the projections, it is also important to sort out the economic rationale behind the relationships between different scenarios. Our base case, to be used as a norm of comparison for projections with other scenarios, is presented in the second subsection below. The alternatives are presented in the third subsection.

3.1 Sectoral Classification

Because of computational considerations and data availability, the number of sectors is restricted to 23. The analytical focus on the impact that changing comparative advantages will have on Swedish economic development suggests more detail in the industrial breakdown than in the corresponding breakdown of the nontradable-goods sectors. Consequently, 15 industrial sectors are given a separate treatment in the model. The classification of these sectors is based on the expected origins of future changes in comparative advantages.

In the model there are three explicitly treated causes for changes in comparative advantages that are related to supply. The first two are differential growth rates of primary factors (capital and labor) and pure technical change (within and outside the industrial sector). A third partially independent determinant originates in the specification of the production functions. Thus technical change is neutral only with respect to the use of capital versus labor but is "primary factor saving" in terms of the relative use of intermediate factors of production.

Causes related to demand that have altered comparative advantages are introduced through the impact of differential growth rates of world markets and changes in world market price structures.

According to these determinants, both demand and supply characteristics of industrial products should influence the aggregation principles. Here supply characteristics are given priority in most instances. In addition, earlier studies of Sweden's changing international specialization (Ohlsson 1977 and Chapters 6, 7, and 10 in 1980b), as well as the nature of the world market scenarios, indicate that the development of human capital or skills also has an important role in this context. Since that factor could not be explicitly incorporated in the production functions, it was instead taken into account in the classification of sectors. In special cases, backward and forward linkages due to transportation costs or technical integration have influenced the sectoral definitions.

Instead of strictly applying a single aggregation principle, we tried to take all these considerations into account in accordance with our best judgment. The following presentation of the sectors provides information about how various factors affected the sectoral classification. The sectors are all listed in Table 1. The table is organized so that the primary sectors (and those strongly related) appear first, followed by the secondary and tertiary sectors.

The energy sector comprises not only all kinds of energy production but also petroleum refineries and asphalt, coal, and oil industries. There is one pure primary sector; mining and quarrying (sector 4). This sector has been a large Swedish export sector for centuries, producing a relatively homogeneous output. Thus it almost exclusively produces iron and pellets of iron rather than more highly priced minerals. Consequently, aggregation causes no particular problems.

There are two mixed primary–secondary sectors: the agriculture, fishery, and basic food sector (number 2) and the forestry, wood, pulp, and paper sector (number 3). Obviously, one of the principles for aggregation has been the strong input–output relationship between primary and secondary production. Moreover, there are so-called economies of integration between them, which in the case of the "agri-food" sector are attributable to transportation costs and policy-imposed ties.* In the case of the forest based sector,

*The agricultural sector is to a high degree excluded from foreign competition in Sweden. Moreover, there is a subsidy system for the basic food industry, which compensates for the otherwise too high input prices created by the agricultural policy. Finally, much of the ownership of the basic food sector is in the hands of farmer cooperatives, which in fact suggests the existence of monopolistic or oligopolistic competition.

TABLE 1 The sectors in the projection analysis.

Number	Production Sector
1	Energy[a]
2	Agriculture, fishing, basic foods
3	Forestry, wood, pulp, paper
4	Mining and quarrying
5	Other foods, beverages, liquor, tobacco
6	Textile, clothing, leather
7	Paper products
8	Chemical products[b]
9	Non-metallic mineral products except petroleum and coal
10	Metals
11	Fabricated metal products
12	Non-electrical machinery, instruments, photographical and optical equipment, watches
13	Transport equipment except ships and boats
14	Electro-technical products
15	Ships and boats
16	Printing and miscellaneous products
17	Hotel and restaurant services, repairs, letting of premises other than dwellings, private services other than banks, insurance, business services
18	Construction
19	Wholesale and retail trade, communications
20	Transport and storage
21	Financial and insurance services
22	Housing services
23	Public services
	Capital goods[c]

[a]Including petroleum refineries and asphalt and coal products.
[b]Excluding petroleum refineries and asphalt and coal products.
[c]The capital goods sector is not a production sector but a "bookkeeping" sector, which aggregates different kinds of capital goods (primarily machinery and buildings) in fixed proportions to an aggregate capital good used in all "real" production sectors.

transportation costs and technical integration economies motivate the aggregation into one sector. The forest based sector is strongly export oriented; the agri-food sector is sheltered from international competition by policy measures.

Apart from these characteristics the primary and primary based sectors also have high or extremely high capital and energy intensities in common. In addition they are all producing relatively standardized products that, with the exception of the products of the agri-food sector, are sold in internationally competitive markets.

There are four semi-raw material based sectors, of which one is foreign trade-exposed: other food, beverages, liquor, and tobacco industries (number 5). The backward linkages of this latter sector are less strong than those of basic food production relative to agriculture. Moreover, the trade-exposed sector 5 is not based as much on domestic raw materials.

Another semi-raw material based sector is the industry for non-metallic mineral products (number 9), which excludes petroleum and coal products. This industry is in part a foreign trade-sheltered sector, particularly because of high costs of transportation.

The remaining two industries within this category are the chemical (number 8) and the metal (number 10) industries. Both contain large parts that have earlier been characterized by more pronounced backward linkages than those appearing to prevail nowadays. It would, however, have been more satisfactory to divide both sectors into at least two parts, one of which would then have been producing the more highly manufactured products. Unfortunately, the present data base did not allow such a breakdown.

Except for one industry, the remaining eight industries (6, 7, and 11–16) are clearly so-called footloose industries; they both are foreign trade-exposed and are little dependent on the location of raw materials production. Three of the seven footloose industries are labor intensive in their production methods: the textile, clothing, and leather industry (number 6), the fabricated metal products industry (number 11), and the electrotechnical industry (number 14).

In many product fields of the first of these footloose industries, the high market shares of less-developed countries (LDCs) suggest the emergence of a price leadership position of low wage countries. The other two sectors have segments in which LDCs have already acquired a substantial competitiveness, but their overall market share is still not high. (See, for instance, OECD 1979, and references and the analysis in Ohlsson 1980c.) The fabricated metal products industry has, for instance, subindustries, that are intensively using semi-modern manual skills and to some extent also technical personnel. Finally, the electro-technical industry contains parts that are among the most technical personnel intensive in relatively "young" technology fields. In other words, these two industries should ideally have been broken down into two or more sectors.

Three of the remaining five industries (7, 12, 13, 15, and 16) have somewhat higher capital intensities. They are primarily distinguished from other footloose industries because of their high human skill intensities (technical personnel and skilled manual workers). The latter feature is most pronounced for the machinery industry (number 12) and also for the transport equipment industry (number 13). Ships and boats (number 15) require less human skill. This industry is at present a government-regulated industry across the world, a characteristic that also holds for the aircraft producing part of the transport equipment industry.

The paper products sector (number 7) was rather dynamic in the 1960s and 1970s with respect to the growth rate of domestic demand. It has an intermediate position on three of the factor intensities discussed above, i.e., on capital, technical personnel, and skilled manual worker intensities. Finally, the miscellaneous industrial production sector (number 16) also includes the printing industry, which has been exposed to a measurable degree of international competition only in the past five years.

All the remaining sectors belong to the tertiary sector, except for capital goods, which was constructed for "bookkeeping" purposes (see footnote *c* of Table 1). Given the focus of the study, we abstain from commenting on these more trade-sheltered sectors.

In summary, the sectoral breakdown is not exactly the most desirable one. It incorporates, however, certain basic technology differences that can be associated with changing comparative advantages. Additional information about the possible sectoral developments can only be introduced in the projections through adjustments of the sectoral values of exogenous variables and parameters. The next two subsections outline the scenarios for these variables and parameters.

3.2 Base Case Scenario

The projections of the model are made for the relatively long periods 1976–1990 and 1991–2000. Our base year is 1975, the last year from which a complete data base is obtainable. With such long time horizons, it is impossible to claim that a particular projection is the most likely one. Instead it is more useful, in terms of policy implications analysis, to establish alternative scenarios in order to find a possible range of structural adjustment and growth paths. The analytical philosophy behind the alternatives can be described as follows.

As mentioned in section 1, there are external and internal causes that change comparative advantages. The main differences between the two, for a small, open economy, are that (a) the external causes can affect the internal ones but the opposite direction of influence can be ignored, and (b) the causes that are controllable for domestic economic policies are all internal. This latter distinction suggests that the policy strategy analysis can be incorporated in the model projections through variations in the values of exogenous variables that belong to the internal cause category.

There are two ways of incorporating changes in comparative advantages through changes in the numerical values of parameters or exogenous variables. One is to change individual sectoral values and the other is to change uniformly all values across sectors. Both ways may have macro as well as structural impacts, but there is one major difference in that the latter, "magnitude" change, does not alter the sectoral comparative advantage ranking, changing only the strength of advantages and disadvantages.

The most obvious example of this is a more rapid accumulation of capital than of labor, which, *ceteris paribus*, strengthens the comparative advantage of capital intensive industries. Indirectly, other magnitude changes, such as the overall rate of world demand growth or of technical change, may also have similar consequences.

Against this background, it was regarded as natural to construct a base case, which combined certain world trade scenarios with those of internal reasons for changing comparative advantages based on the official Swedish long-term forecasts. This means, in turn, that the "domestic scenarios" in the base case more or less project the future causes in comparative advantage changes to be similar in magnitude and structure to those of the past two decades.

As is clear from section 2, the world market scenarios consist of assumptions about growth rates of the world market for trade-exposed sectors, and changes in world market relative prices. The most globally comprehensive and consistent set of estimates of the two sets of variables is found in Leontief (1977) in Scenario A, which is the most "endogenous" of that study. Except for a few regions, neither the gross domestic product (GDP) nor employment are assumed to attain target values. Instead those magnitudes are endogenously determined under the constraints incorporated in the global model system utilized in the study.

The world market price assumptions are based on projections of production costs in the economy of the United States. Implicitly, therefore, it seems to presuppose that US producers are able to maintain much of the same price leadership role in the world economy as they had in the 1950s and 1960s. Although the European and Japanese challenges altered this role in the ten years before our base year and the industrialization of LDCs is about to alter it in one or two sectors, this basic assumption will not be questioned in the present study. The issue, however, is important enough to be a topic for another report. For the sake of brevity it is not treated here.

This limitation on the realism of our world market scenarios is perhaps not as serious as it might first appear. The reason is that the use of historical data on US production costs for projections of world market relative price changes is also possible in another case. Suppose that US industry acts as a price taker on the world market but as a consequence of its size has no factor-biased intra-industry specialization. Then its domestic prices and costs of production follow those established by the world market.

The second set of world market variables obtained from the same source is the growth rates of world market by commodity groups. There is not much to say *a priori* about these figures in terms of their theoretical or empirical underpinnings. Both sets of variables are presented in Table 2 together with some other scenario variables.

If the cross-sectoral differences in the two sets of world market variables are evaluated, however, two rather surprising changes compared with historical experience should be noted. One is the extremely favorable development for exporters of textiles, clothing, and leather with respect to both the relative price change (a moderate decrease) and the world market growth rate. This sector and three others have the most dynamic growth rate: paper products, non-metallic mineral products,* and printing and miscellaneous products.

Given the above-mentioned nature of the relative price forecasts, it appears that Leontief's price forecast may be subject to a bias from an intra-industry specialization in the US on less price sensitive segments of the textile and clothing sector. Thus for this particular sector we consider the price leadership role of the US economy and the assumption that no factor-biased intra-industry specialization is unrealistic. This may follow as a consequence of successful LDC market penetration. The associated relative cost increases in the US industry have then a built-in upward bias if taken as a projection of the world market relative price. In turn, this may explain the rather high projected world market growth rates for these products. For this reason the projections of the Swedish textile industry must be considered to be rather optimistic both from the price and the world market growth points of view.

Another remarkable projected change is the comparatively low market growth figures for certain engineering sectors (non-electrical machinery, transport equipment, and electro-technical products) and the chemical sector compared with both shipyards and certain raw material based sectors (sectors 3, 9, and 10).

With these two projected changes in mind, it may be concluded that Leontief's study has used a constellation of assumptions that is very favorable for an industrial composition of a typical developing country. Consequently, the world market scenarios utilized in the present study must be interpreted as being on the pessimistic side for Sweden's high skill intensive, footloose industries and overly optimistic for its raw material, raw material based, and raw labor intensive footloose sectors. Accordingly, the projected structural adjustments must be considered to be smaller than expected from the history of the first five years of the projection period.

*This sector also appears to obtain a remarkably favorable world market future, although this judgment is based more on the composition of the domestic industry than on past trends in world trade.

TABLE 2 Sectoral specifications of world market scenarios, price elasticities, and productivity growth.

Sector number	Production sector	Percentage growth in world trade		Relative price in the year		Import price elasticity	Export price elasticity	Yearly rate of productivity growth in percent
		1975–90	1990–2000	1990	2000			
1	Energy	–	..	2.71	3.05	–	–	1.0
2	Agriculture, fishing, basic foods	0.0	1.0	1.07	1.11	1.5	–2.5	1.0
3	Forestry, wood, pulp, paper	7.0	6.0	0.91	0.90	0.8	–1.5	1.0
4	Mining and quarrying	4.0	4.0	1.00	1.00	1.0	–2.0	1.0
5	Other foods, beverages, etc.	1.0	1.0	0.95	0.93	1.0	–2.0	1.0
6	Textile, clothing, leather	8.0	7.0	0.93	0.92	1.5	–3.0	2.0
7	Paper products	8.0	7.0	0.87	0.86	0.3	–0.6	2.0
8	Chemical products	6.0	5.0	0.98	0.99	1.0	–1.5	4.0
9	Non-metallic mineral products	8.0	7.0	0.93	0.94	0.5	–1.0	1.0
10	Metals	4.0	3.0	0.97	0.96	0.8	–1.5	3.0
11	Fabricated metal products	4.0	3.0	0.97	0.96	1.5	–2.5	2.0
12	Non-electrical machinery, etc.	6.0	6.0	1.00	1.00	1.8	–2.5	2.0
		(5.0)[a]	(5.0)	(0.89)	(0.89)			
13	Transport equipment	6.0	5.0	0.95	0.94	0.6	–1.0	2.0
14	Electro-technical products	7.0	6.0	0.90	0.93	0.8	–1.2	2.0
15	Ships and boats	5.0	5.0	0.85	0.82	1.0	–1.5	2.0
16	Printing and miscellaneous products	8.0	7.0	0.87	0.86	0.8	–1.2	2.0
17	Hotels, restaurants, etc.	4.0	4.0	1.00	1.00	0.2	–0.3	0.5
18	Construction	4.0	4.0	1.00	1.00	..	..	1.5
19	Wholesale and retail trade, etc.	4.0	4.0	0.91	0.91	0.2	–0.3	1.5
20	Transport and storage	5.0	4.0	0.95	0.96	0.2	–0.3	1.5
21	Financial and insurance services	4.0	3.0	1.01	1.00	0.2	–0.3	0.5
22	Housing services	–	–	–	–	–	–	1.0
23	Public services	–	–	–	–	–	–	0.0
24	Capital goods	–	–	..	–	–	–	..

[a]The figures in parenthesis are from Leontief (1977).

Moreover, the same conclusion holds for any country as far as the structural influence of changing relative prices is concerned because of the rather small spread in projected prices within the industrial sector. The only exception to this latter observation is the energy sector, where the relative price level more than triples compared with all other sectors.

As can be seen from Table 2 we have adjusted the market growth rate from 5 to 6 percent and assumed a more favorable relative price development for the non-electrical machinery sector. It is not the above-noted possibility of changes in the US price leadership role that motivates the adjustments in this case. Instead, it is the Swedish intra-industry specialization in investment goods for raw material and raw material based production, etc., that constitutes the basis for these adjusted figures. According to the Leontief projections the rapid growth of these latter sectors should be associated with a more than average rate of increase in their demand for investment goods. Moreover, the production of such heavy machinery has had a lower rate of technical change than, for instance, computer and office machinery production, which is also part of the non-electrical machinery sector. For this reason the relative price decrease of the cited study appears to be biased downward for a machinery sector with the present Swedish output mix.

Table 2 also provides the sectoral relative price elasticities of imports and exports and the annual rates of productivity growth. The former two sets of figures have been chosen on the basis of estimates in Hamilton (1979) on import share relative price elasticities for the period 1960 1975. Generally speaking, the price elasticities of this study seem to be rather low. Combined with the small relative price changes, this is likely to produce a rather low impact on structural change.

The price elasticities estimated by Hamilton were changed for only three sectors: chemical, non-electrical machinery, and transport equipment. The elasticities were adjusted downward for the first two and upward for the last sector. The assumed high elasticities for chemicals and non-electrical machinery are probably due to the combination of low tariff barriers and rapid intra-industry trade and specialization in the 1960s and 1970s rather than particularly high substitutability with similar products produced in other countries. Similarly, the estimates of the transport equipment industry are presumed to be low because of the development of favorable relative tariff rates (see Ohlsson 1980b, chapter 6).

The import price elasticities have the same rank ordering as the export price, but lower absolute values. This is attributed to proximity advantages in the home market for domestic producers. Since Sweden is geographically rather isolated from its main foreign markets and because of the large surface over which the economy is spread, the differences between exports and imports are usually large in absolute terms. Small relative differences were introduced for homogeneous industries with highly tradable products. Needless to say, these differences introduce a stronger element of arbitrariness for export price elasticities than for the import price elasticities.

Finally, the assumed annual growth rates of productivity presented in Table 2 are based on projections by the Swedish Ministry of Economic Affairs (see Restad 1976). These projections have since been revised downward. The revised values, however, were unavailable to us in some of the more detailed sectors. For these sectors we made proportional downward revisions. The forestry, wood, pulp, and paper sector has been attributed an even lower figure. This is because the decreasing availability of domestic raw material supplies is assumed to increase the costs of additional supplies.

In accordance with the figures obtained from the Ministry of Economic Affairs we have assumed a yearly increase of 1.8 percent in real public consumption throughout the period 1975–2000. The corresponding figure for the real capital stock of the economy is set at 2.5 percent per year. Labor supply measured in man-hours is assumed to remain constant at the 1975 level. This last assumption allows for the fulfillment of ambitious goals about increased labor participation rates in an almost stable Swedish population, mainly through an enhanced degree of part-time work. Consequently, the differential growth rates for the two primary factors induce, *ceteris paribus*, a more capital intensive specialization.

This concludes our presentation of the base case assumptions. The principles and figures for the alternative scenarios are discussed next.

3.3 Alternative Scenarios

Early computations suggested that macroeconomic development and the sectoral distribution of employment were rather insensitive to reasonable changes in relative prices *or* price elasticities. In order to alter the results substantially, the magnitudes on both had to be altered considerably. Instead the projections turned out to be more sensitive to changes in rates of world market growth and domestic productivity. For this reason, the alternative scenarios are built on alternative assumptions about the latter two sets of exogenous variables.

The simplest change is to alter the magnitudes across all sectors and not the sectoral differences in world market growth rates and productivity rates. It is reasonable to adjust the magnitudes downward by 1 percent per annum for all tradable sectors, i.e., to let the world market growth rate be even lower than was projected in Leontief (1977). Given the historically low rates of productivity growth, the 1 percent change in productivity rates results in an upward change. Even so, the rate of productivity growth falls below that of the 1960s. Calling the base case number I, three alternative combinations of assumptions are used:

- Case II the same as the base case in all respects except for a 1 percent higher annual productivity growth rate in all sectors
- Case III the same as the base case in all respects except for a 1 percent lower rate of world market growth in all tradable sectors
- Case IV combines the two adjustments of cases II and III, i.e., compared with the base case both a 1 percent higher general, annual productivity growth rate and a 1 percent lower general rate of world market growth.

Apart from these cases, the sensitivity of certain macroeconomic results to alternative assumptions concerning capital accumulation and labor supply is also analyzed. For simplification these alternative assumptions have been condensed and are not discussed in detail.

4 PROJECTIONS

The results of the model simulations are given in the following subsections. In subsections 4.1 and 4.2, base case results are presented for the projected macroeconomic development and sectoral development, respectively. Subsection 4.3 deals with the consequences of altered world market and productivity assumptions at the macroeconomic level, whereas the ensuing subsection deals with the corresponding sectoral consequences. In order to avoid repetition and to acquire a better tie to the subsequent analysis of regional implications in section 5, the sectoral consequences are described in terms of employment consequences.

4.1 Macroeconomic Developments: The Base Case

The model was solved for the years 1990 and 2000, but in most cases we prefer to present the macroeconomic results in terms of annual percentage rates of change during the periods 1976–1990 and 1991 2000. It was assumed that the intersectoral profit differences prevailing initially will be eliminated by 1990. Consequently, the first of these subperiods can be regarded as a period of adjustment, both from a disequilibrium to an equilibrium state of the economy and to certain exogenous changes inside and outside the economy.

To begin with, we focus on the projected development of GDP, aggregate real consumption, industrial production and employment, the functional distribution of income, and relative size of the public sector.

Table 3 contains the projected growth rates for real GDP and aggregate private consumption during the two subperiods 1976–1990 and 1991–2000. These data contain three striking results: the rate of economic growth is considerably lower than the postwar average, the two subperiods are different, and finally, the share of private consumption in the gross national product (GNP) increases over the whole period. In what follows, possible explanations of these three results are offered.

TABLE 3 Projected annual growth rates for real GDP and aggregate private consumption, 1976–2000.

	Projected growth rates in percent	
Variables	1976–1990	1991–2000
GDP	2.2	1.8
Private consumption	3.0	2.6

During the period 1950–1975, the average rate of economic growth (growth of GDP) in Sweden was 3.6 percent per annum. If the "bad" years in the beginning of the 1970s are excluded, the average rate for 1950–1970 becomes 3.8 percent per annum. This means that, according to our projections, Sweden has entered a period with considerably slower economic growth than was experienced during the earlier postwar period.

There are many factors behind this development: slower rate of capital formation and technical change, stagnation in the supply of labor* (in man-hours), and a relatively

*Observe that the labor force is assumed to be fully employed in all model simulations.

fast growth of an already large public sector, which, in accordance with national accounting conventions, is here attributed a zero productivity increase. In addition, some private service sectors, with a relatively slow rate of productivity increase, grow faster than GDP.

The second startling feature of our results is the difference between the two subperiods; the rate of growth is considerably higher from 1976 to 1990 than from 1991 to 2000. The explanation is simple and straightforward. The initial year, 1975, shows many features of a disequilibrium situation. The average rate of profit was very low and the intersectoral differences in terms of profit rates were significant. In two of the 23 aggregated sectors, losses were revealed by the data. Thus a sectoral reallocation of resources could produce substantial efficiency gains. This is exactly what happens between 1975 and 1990 in our projection.* Net investments are concentrated in a few relatively profitable sectors, and old capital is not replaced in some sectors. This development tends to equalize profit rates and thus the marginal productivity of capital in the different sectors. This equalization leads to an increase in the average productivity of the economy's resources. During the second subperiod, however, these potential reallocation gains are already exploited, and capital accumulation and technical change are the main sources of economic growth besides the reallocation gains associated with changing world market prices.

With this background even the low growth rates displayed in Table 3 might be too optimistic in practice. In a process where efficiency in resource allocation is a significant source of economic growth, labor and capital markets have to function quite smoothly; without much delay, resources have to be reallocated from stagnating to expanding sectors. The present institutional framework of the Swedish economy does not seem to be well-suited for fostering such a process. In particular, the interregional and intersectoral labor mobility may be substantially lower in the future than in the 1950s and 1960s. This might be a result of changes in the institutional framework of the labor market in the 1970s and the implementation of very ambitious policy goals aimed at stabilizing employment on the regional or county, and sometimes even the firm, level.

As mentioned in section 3.2, one factor that suggests growth rates are too low is the relatively small amount of incentives to structural adjustment hidden in the Leontief (1977) world economy projections. This reduces the intersectoral differences in terms of comparative advantage changes and thus the contribution to economic growth from intersectoral reallocation of resources.

Another feature of our 1976–1990 projection is that the profit level in the private sector of the economy, measured as total *pre-tax* net profits in relation to the replacement value of the capital stock, increases from 3.8 to 4.7 percent. This increase contributes to the growing share of capital income in total national income. It can be questioned whether such a development would be politically accepted in Sweden without a negotiated change in the distribution of ownership in the industrial sectors.

This is a very crude way of posing the income-distribution problem, however; the marginal productivity of capital need not be equal to the *after tax* income from capital. The critical point of the analysis is therefore whether the rate of profit after taxes is high

*In Bergman and Pór (1980) the potential reallocation gains are estimated, using the same model and data base. The results indicate that full exploitation of the potential reallocation gains in 1975 would lead to a GDP that would be 4 percent higher than the actual value.

enough to bring about the assumed annual 2.5 percent increase in the economy's stock of capital.

The third striking result is the relatively fast growth of private consumption. (As will be discussed in some detail in section 5, this result does not conform to the long-term projections carried out by the Ministry of Economic Affairs.) By assumption, investments grow by 2.5 percent per annum and real public consumption by 1.8 percent per annum. Since GDP grows by an average of 2.0 percent per annum, an average rate of private consumption growth of 2.8 percent per annum implies that exports grow slower than GDP. This is exactly what takes place in our base case projection. Due to a significant terms-of-trade improvement (1.9 percent per annum despite increasing real oil prices), external balance is maintained although real exports only grow 1.7 percent per annum.

The terms-of-trade improvement is a consequence of the fast growth of world market trade in relation to Swedish economic growth together with the incorporation of explicit price-dependent export functions in the model. Thus external demand increases will be met by a combination of export supply and export price increases. A projected reallocation of exports toward commodities with relatively increased world market prices has a similar effect on the terms of trade.

From an empirical point of view, however, this result should be interpreted with care. The projection includes a considerable gap, about 40 percent, between Swedish and world market prices for some commodity groups. We have no such experiences from the estimation period, and consequently we do not know whether our estimates of price elasticities in the export and import functions are still valid for the price relations prevailing in our projections for the year 2000.* Another reason for caution when interpreting this result is the rapid net accumulation of foreign debt in Sweden in the past five years, which has led to a new goal for economic policies: the repayment of the outstanding foreign debt in the 1980s. Therefore the current account is targeted to yield a surplus, which cannot be achieved unless, *ceteris paribus*, there is a deterioration of the terms of trade. Finally, the terms-of-trade development projected by the model is sensitive to world market assumptions: slower world market growth worsens the terms-of-trade development.

Table 4 contains some results on the semi-macro level. Industrial production grows slower than GDP and industrial employment decreases during the entire projection period. Energy consumption grows considerably slower than the 5.5 percent per annum experienced during the period 1950–1972. A few comments should be made about these results.

During the postwar period, industrial production has, in general, been growing faster than GDP in Sweden. According to our projection, the reversed relation would hold in the future. The consumption of industrial goods, however, continues to grow faster than GDP. Thus the basic difference is that the import share in the domestic supply of industrial goods increases considerably: from 27.8 percent in 1975 to 40.2 percent in

*Section 6 gives a critical appraisal of this approach. Chapters 5 and 7 of Ohlsson (1980b) show considerable differences between unit prices of exports and imports at a detailed level of industrial breakdown compared with that used in the present paper. Intra-industry specialization appears, furthermore, to be characterized by exports of higher priced product variants and imports of lower priced ones compared with other OECD countries. The market share implication of this specialization, however, is not as simple as the one used above.

TABLE 4 Projected annual growth rates for industrial production and employment and total energy consumption, 1975–2000.

Variables	Projected growth rates in percent	
	1975–1990	1991–2000
Industrial production	1.9	1.5
Industrial employment	–1.0	–2.3
Total energy consumption	1.1	2.2

2000. This is, of course, the mirror image of the above-mentioned terms-of-trade improvement and slow export expansion. The much slower growth of exports and production for the domestic market explains, in turn, why industrial employment decreases at a fast rate. By the turn of the century, the industrial sector would then have lost about 30 percent of its 1975 employment (in man-hours) to primarily service-producing sectors. Another way of expressing the causes behind this development is to say that the industrial sector is squeezed between competition with foreign producers in commodity markets and foreign trade-sheltered producers (particularly the public sector) in the (primary) factor markets. The latter is the result of the absence of (or low) productivity growth rates in tertiary sectors and the lack of strong demand-restricting factors when production costs increase.

The relatively slow growth in the rate of energy consumption is, of course, partly a result of the slow growth of industrial production. It is also, however, a result of substitutions of capital and labor for energy, induced by an increasing relative price of energy. Between 1950 and 1972, the real price of energy decreased by nearly 3 percent per annum. In our projection the average rate of increase between 1975 and 2000 is 1.0 percent per annum. Most of the price increase, however, takes place during the first subperiod, primarily as a result of oil price increases but also as a result of the rate of interest increase, which affects the capital intensive energy sector more than other sectors. The uneven development of the relative price of energy explains the differences in energy consumption growth between the two subperiods.

On *a priori* grounds, it cannot be ruled out that the projected slow growth of industrial production in the Swedish economy is the result of increasing energy costs, but a closer look at the results does not support such a hypothesis. The share of energy costs in total production costs is generally low in the industrial sectors, between 5 and 10 percent at the terminal point (the year 2000) compared with 3 and 8 percent in 1975. This means that the projected energy price increase still has a relatively minor impact on the development of production costs in industrial sectors.

Moreover, as long as Swedish energy prices change in the same way as energy prices in other countries, the development of Sweden's comparative advantages should not be affected much by increasing relative prices of energy. To put it another way, the tripling of world market energy prices should also be reflected in Leontief's estimates of the world market prices for sectors requiring energy. In the base case projection, we have assumed an "unchanged energy policy" in Sweden; that is, we have not assumed any major changes in production technology in the energy sector or in the taxation of energy. The world market price projections, obtained from the Leontief study, rest on similar assumptions.

During the 1970s, a conflict arose between private and public consumption. In accordance with the projections obtained from the Ministry of Economic Affairs, we have assumed that real public consumption will increase by 1.8 percent per annum between 1976 and 2000. In our projection, this leads to an increase in public employment of 1.8 percent per annum. As a result, the share of the labor force employed by the public sector increases from 22.6 to 36.9 percent. The price index for public consumption increases by 2.2 percent per annum in relation to the general price level. Thus, in our projection, the share of public consumption expenditures* in the nominal national income increases from 26.8 percent to 36.9 percent in 2000. The impact of this development on the share of private consumption expenditures is somewhat mitigated by an annual 0.6 percent decrease in the relative price of capital goods, which in conjunction with fixed development of real investment expenditures leads to a gradual decrease of the gross savings ratio. As can be seen in Table 5, however, the projected development implies a slow growth of disposable income for the household sector.

TABLE 5 Aggregate demand categories as a percentage share of GDP in constant and current prices.

	Constant prices		Current prices	
Demand categories	1975	2000	1975	2000
Private consumption	51.8	64.6	51.8	44.7
Public consumption	26.8	25.4	26.8	36.9
Gross investments	22.3	25.7	22.3	18.5
Net exports	–0.9	–15.6	–0.9	0.0

To sum up, the projection based on base case assumptions implies a considerably slower rate of economic growth in Sweden in the future than during the first postwar decades. Moreover, there is a significant shift of demand and reallocation of resources from the industrial sector to the service sector.

4.2 Projected Sectoral Developments: The Base Case

Slow growth of the industrial sector as a whole does not prevent a substantial variation among industrial sectors. This can be seen in Table 6. The figures can be compared with the annual growth rate of GDP, which amounts to 2 percent for the whole 25-year period. As many as seven of the industrial sectors have higher projected growth rates than 2 percent; the most outstanding ones are paper products and electro-technical products. Apart from the latter industry, however. the growth rates of the engineering sectors (11–15), which are the growth sectors, are unfavorable considering the expectations in Sweden, as well as in other industrial countries. A rapid decline of the ships and boats sector is expected and after five years has already been partially fulfilled, despite the rapid world

*The share of transfer payments in nominal national income is presently about 30 percent.

TABLE 6 Projected annual growth rates of real production and of employment by sector 1975–2000.

	Projected growth rates in percent	
Production sector	Production	Employment
Energy	1.8	–3.2
Agriculture, fishing, basic foods	2.1	–1.7
Forestry, wood, pulp, and paper	1.9	–0.1
Mining and quarrying	–0.2	–3.3
Other foods, beverages, etc.	2.1	–0.7
Textile, clothing, leather	0.8	–2.4
Paper products	4.7	0.2
Chemical products	2.2	–3.2
Non-metallic mineral products	2.1	–0.7
Metals	–0.4	–5.1
Fabricated metal products	0.0	–3.0
Non-electrical machinery, etc.	0.8	–2.3
Transport equipment	1.1	–1.8
Electro-technical products	2.5	–2.6
Ships and boats	–1.9	–5.0
Printing and miscellaneous products	2.1	–1.0
Hotels, restaurants, etc.	2.1	0.5
Construction	2.4	0.4
Wholesale and retail trade, etc.	1.7	–1.1
Transport and storage	1.9	–0.8
Financial and insurance services	1.9	0.8
Housing services	2.7	–2.6
Public services	1.8[a]	1.8[b]

[a]Assumed to be exogenously given.
[b]Follows from assumptions of zero rate of productivity change and no possibilities of factor substitution.

market growth rate. Consequently, it is the combination of bleak relative price developments and moderate productivity increases that explain this result.

Despite the absence of powerful external incentives for structural change embedded in the world market scenarios based on Leontief, the typical stagnant industries are those that were recognized as such in the later 1970s. Along with the ships and boats sector mentioned above, we can expect negative growth rates for the mining and quarrying industry and the metals industry. The forestry, wood, pulp, and paper industry continues to have a relatively good growth performance, a result which appears attributable to Leontief's high world trade projections as well as to rapidly expanding deliveries to the most spectacular growth sector: the paper products industry.

In summary, the structural adjustments within the industrial sector appear to continue with regard to stagnating industries, but the trends from the 1960s and 1970s for some of the expected Swedish future growth industries are altered. This is especially the case for the non-electrical machinery industry. It is the combination of rather "pessimistic" world market scenarios for these industries and possibly the projected competitive domestic market for primary factors of production (especially from service sectors), that are probably accounting for this bleak outcome. Consequently, the small external incentives for structural change reduce the growth of the likely expansive sectors but do not protect the problem sectors from stagnation or contraction. This result explains the poor outlook for industrial employment. Even at the assumed historically low rates of productivity increases, the industrial sectors cannot maintain their employment levels, except in the expansive paper products industry.

In the following section we dwell upon this issue in more depth. Let us only direct attention here to the discussion in the preceding section about the terms of trade increase and the related slow growth of real exports compared with real imports, industrial production, and GDP. These features would mark the ending of a long historical record of export-led growth; Sweden would lose market shares rapidly, domestically as well as abroad.

4.3 Macroeconomic Developments: Alternative Cases

At this point in the analysis of the projections, we have obtained a fairly evident perception of the main causes behind economic development at large: reduced domestic sources of economic growth, smaller than expected external incentives for intersectoral structural adjustments in the trade-exposed sector of the economy, and rapidly growing world markets. It should also be clear by now why the alternative assumptions of cases II–IV were chosen using increased productivity growth rates and decreased rates of world market growth; both influence the industrial sector in the same way, by reducing the pressures incurred through the improvement in Sweden's terms of trade. Thus we alter two of the three major growth pattern determinants mentioned above, but keep the third (i.e., the incentives for structural change between industries) fundamentally unchanged.

Table 7 summarizes the projected development of the aggregate demand components and the terms of trade between 1976 and 2000 in the base case and the three other cases described in subsection 3.3. The results in Table 7 clearly indicate that the projected rates of change of the macro variables are quite sensitive to variations in productivity and world market assumptions. Although the variations made in these assumptions are arbitrary, they are well within the range given by the uncertainty of the long-term projections utilized in the construction of the scenarios. The results indicate that the uncertainty in

TABLE 7 Projected annual growth rates 1976–2000 for selected macroeconomic variables.

	Projected growth rates in percent			
Variables	Case 1	Case II	Case III	Case IV
Private consumption[a]	2.9	4.0	2.2	3.6
Public consumption[a]	1.8	1.8	1.8	1.8
Gross investment[a]	2.6	2.6	2.6	2.6
Exports[a]	1.7	3.4	2.0	3.7
Imports[a]	3.4	3.7	2.6	3.0
GDP[a]	2.0	3.2	2.0	3.2
Terms of trade	1.9	0.4	0.7	-0.6

[a]In constant (1975) prices.

these exogenous conditions leads to a significant uncertainty in the long-term projections of GDP, real consumption, and other macroeconomic variables.

One of the most interesting results obtained from these experiments is the remarkable difference the variations of underlying assumptions made in terms of changes in the export growth rate. According to Table 7 the rate of export growth is mainly determined by the productivity increase (compare cases I and II with cases II and III, respectively). Observe here also that even this higher productivity growth rate falls below the earlier postwar experience.

In summary, it is quite likely that the contributions to economic growth of the overall productivity change are lowered in comparison with the contributions from factor accumulation in two ways: low sectoral productivity growth rates and small external changes in comparative advantages. In this respect future economic development would substantially deviate from past records. As has been shown by Åberg (1969) and in the updated figures in IVA and IUI (1979), the percentage contribution of the so-called technique factor has increased over the postwar period at the expense of the contributions of capital and labor accumulation.

This shift in the role of factor accumulation is not at all a consequence of higher accumulation rates. On the contrary, both primary factors of production increased more in supply before the projection period than during it. Against this background it is interesting to investigate the sensitivity of the projections with respect to the supply of capital and labor. Such a sensitivity analysis for the results in the year 2000 can be easily revealed in the form of elasticities of endogenous variables with respect to the total supply of capital and labor (base case assumptions). The main findings are summarized in Table 8. The elasticities are valid within a range of ±10 percent for variations of the exogenous variables in question.

TABLE 8 The calculated elasticity of GDP and real private consumption with respect to selected exogenous variables.

	Elasticity	
Selected exogenous variables	GDP	Real private consumption
Total supply of capital	0.33	0.35
Total supply of labor	0.74	0.83

Again the projections turn out to be quite sensitive to assumptions about exogenous conditions. Apparently the conclusion that the Swedish economy has entered a period with a *significantly* slower rate of economic growth than during the earlier postwar decades holds only under scenario definitions I and III but not with more normal rates of technical progress and higher capital and labor accumulation rates. In all projection cases, however, the rate of GDP growth is lower than the 3.6 percent per annum during the period 1950–1975.

Another important result obtained under base case conditions is that industrial production is projected to grow more slowly than GDP in the future. This result, which represents a change in postwar trends, holds in all cases except case IV where industrial production grows by 3.4 percent per annum and GDP by 3.2 percent per annum. In all cases, however, total employment (in man-hours) in the industrial sectors declines by more than 1 percent per annum. The overall impression given by the table is that the best results for GDP and private consumption growth would be achieved if the supply of labor could be increased. It can only be substituted for with a more than double rate of increase in capital productivity.

4.4 Projected Sectoral Developments: Alternative Cases

As mentioned in the introductory part of this section the sectoral implications of the four cases will be analyzed in terms of employment composition changes. The intersectoral variation is not much affected by variations in the rate of productivity and world market increases. In addition, a study of compositional changes in employment puts more of the results in a policy perspective because of the priority of various employment goals in Sweden. The full employment equilibria projected here, however, do not allow an analysis of the full employment goal.

Table 9 presents the sectoral breakdown of employment in 1975 as well as in the year 2000 according to the four alternative cases. Let us first concentrate our attention on the broad changes in the employment composition.

The tertiary sector contributed to more than 60 percent of the national employment in 1975. About 25 percent of the labor force was occupied in the production of public services. The base case projects the tertiary employment share to 76 percent in the year 2000 with 39 percent in the public service sector. A service economy will have arrived, and a large part of it will be organized as public services between privately and publicly produced goods and services, according to the present division of labor in Sweden.

Cases II and III have in common a 1 percent per annum higher productivity growth in all sectors, including the public sector. Evidently, this makes quite a difference in terms of employment shares. Tertiary employment will then only expand from 62 to about 70 percent, mainly because of the much lower rate of growth of employment in the public sector. Its employment share of the whole tertiary sector increases from 40 to 44 percent compared with more than 50 percent in the base case projection.

Accordingly an overall and (in absolute terms) equal rise in the rate of productivity growth improves the employment situation for primary and secondary sectors *vis-à-vis* the tertiary and for private services employment compared with public services employment. Apparently, it is the decline in sectoral differences in the rate of productivity growth that accomplishes this change in our results. The more optimistic the scenario concerning productivity growth in the tertiary sector compared with the commodity producing sectors, the less the employment shift toward more service-producing jobs.

Finally, it is worth noting the changes in the composition of employment between primary and secondary sectors. Table 10 gives a proper overview of the summary figures.

Only the ships and boats sector is excluded from the overview. According to all projections this sector is the most dramatically declining one in terms of employment shares, despite rather optimistic projections of world market growth rates.

TABLE 9 The sectoral contribution to total employment in 1975 and in the year 2000 for cases I–IV.

Production sector	Employment shares in percent				
	1975	Case I 2000	Case II 2000	Case III 2000	Case IV 2000
Energy	1.0	0.5	0.6	0.5	0.4
Agriculture, fishing, basic foods	7.0	4.6	5.9	5.7	4.3
Forestry, wood, pulp, and paper	4.9	4.8	5.8	6.0	5.0
Mining and quarrying	0.5	0.2	0.3	0.4	0.3
Other foods, beverages, etc.	0.8	0.7	0.9	0.8	0.6
Textile, clothing, leather	1.9	1.0	1.5	1.6	1.1
Paper products	1.1	1.2	1.2	1.1	1.1
Chemical products	2.1	0.9	1.2	1.3	1.0
Non-metallic mineral products	1.1	0.9	1.0	1.0	1.0
Metals	2.2	0.6	0.8	0.9	0.7
Fabricated metal products	3.0	1.4	1.9	2.2	1.7
Non-electrical machinery, etc.	4.5	2.5	3.4	3.9	3.0
Transport equipment	2.5	1.6	2.0	2.1	1.7
Electro-technical products	2.4	1.3	1.6	1.6	1.3
Ships and boats	1.2	0.3	0.4	0.4	0.3
Printing and miscellaneous products	2.1	1.6	1.9	1.9	1.6
Hotels, restaurants, etc.	9.7	11.1	12.8	12.2	10.4
Construction	9.9	10.8	9.3	9.3	10.8
Wholesale and retail trade, etc.	7.3	5.6	6.7	6.8	5.7
Transport and storage	5.8	4.7	5.2	4.9	4.3
Financial and insurance services	3.6	4.4	4.6	4.6	4.3
Housing services	0.7	0.4	0.5	0.4	0.3
Public services	24.9	39.0	30.6	30.6	38.9
Total employment	100.0	100.0	100.0	100.0	100.0

TABLE 10 Employment shares in selected sector groups in 1975 and in the year 2000 for cases I and II.

	Employment shares in percent		
Sector groups	1975	Case I 2000	Case II 2000
Primary and raw material based sectors (1, 2, 3, 4)	13.4	10.1	12.6
Semi-raw material based sectors (5, 9, 10)	4.1	2.2	2.7
Raw labor intensive footloose sectors (6, 11, 16)	7.0	4.0	5.3
Paper, chemical, and most engineering products (7, 8, 12, 13, 14)	12.6	7.5	9.4

The projected sectoral employment shares summarized in this way have the same story to tell. The intersectoral changes in primary and secondary sectors are surprisingly small. In fact the employment share decline is considerably smaller in the primary and raw material based sectors than in the remaining categories of secondary sectors. This outcome stands in sharp contrast to historical records for at least the last three or four decades.

Another contrast to past developments is that the chemical and engineering sectors (9–15) have such a mediocre future. As mentioned earlier this result is mainly attributable to the world market scenarios in Leontief (1977). These scenarios do not provide much incentive for structural changes within the industrial sector. In fact it appears as if Leontief's relative price and market growth projections show an opposite tendency for future structural incentives than has been experienced in the last several decades. There is, therefore, good reason to wonder whether these projections are compatible with both our general knowledge about the secular trends and the projected trends in our own model toward a more service producing economy.

5 REGIONAL IMPLICATIONS

The projected full employment equilibria presume smoothly adjusting commodity and factor markets in the 25-year time horizon. Even though the time period is long, there might be adjustment rigidities that are strong or long-standing enough to prevent the projected reallocation of resources from taking place. Such rigidities may be endogenous to the economic system or policy imposed. Compared with several other small, open economies, Sweden differs in its spatial extensiveness; even the industries themselves are spread over most of the country and scattered in many, often relatively dispersed, villages or small towns.

The combination of a small, open, and spatially extensive economy may impose adjustment rigidities in two ways. First, the geographical mobility of factors and products may be more limited than in other small economies. Second, the regional population and employment goals may have a relatively high priority compared with other goals.

This report focuses on the latter type of rigidity. Instead of making quantified projections of the regional developments associated with the projected national–sectoral one, we have settled for a more qualitative approach. By comparing the regional implications of sectoral employment presented in section 4.4, it is possible to draw some general conclusions about the nature of the future regional labor market adjustment problems. The magnitude of the adjustment problems suggests, in turn, whether or not the projected developments are politically feasible in the sense that they could be acceptable with the current goal priorities.

The discussion on this point must combine two regional adjustment problems. One is historically associated with the contraction of the primary sectors in northern Sweden and the other with the rapid metropolitan growth of especially the Stockholm region, which is attributable to the expanding tertiary sector. Both these sectors incorporate many production units that are not as footloose as the corresponding establishments in the manufacturing industry.

The historical concentration of the tertiary sectors in metropolitan Sweden is shown in Table 11. The three metropolitan counties surrounding Stockholm, Gothenburg, and Malmoe in 1975 had about 36 percent of Sweden's total population and 39 percent of

TABLE 11 Population and employment shares (in percent) for three metropolitan counties (Stockholm, Gothenburg and Bohus, and Malmoehus) in 1970 and 1975.

	All three metropolitan counties		The metropolitan county of Stockholm, capital city of Sweden	
Production sector	1970	1975	1970	1975
Agriculture, forestry, fishing	15.6	15.7	3.4	3.4
Mining and quarrying	9.7	5.7	4.3	3.1
Manufacturing	31.0	30.1	13.4	12.5
Electricity, gas, heat, and water production	44.4	39.9	24.4	21.2
Construction industry	35.9	35.2	18.2	17.2
Wholesale and retail trade, hotels, restaurants	47.1	46.3	25.0	24.5
Transport and communications	47.4	48.7	25.5	25.4
Finance, insurance, housing services, consulting	61.1	57.8	40.6	38.2
Public services	42.3	43.0	24.2	24.7
Total employment in above sectors	38.3	38.7	19.9	20.1
Total population	36.1	35.9	18.3	18.2

SOURCE: Table 3.6 in Göteborgs kommun (1978).

its total employment. Their combined share of total employment was substantially higher in each one of the tertiary sectors. This was particularly the case for the finance, insurance, housing services, and consulting sector.* Moreover, most of this location bias was due to the high shares of the capital city of Stockholm.

*The decline in the employment concentration of this sector between 1970 and 1975 is probably due to a decentralization of certain large insurance companies and commercial banks. This decentralization was made possible by the relatively early and rapid introduction of computers and computerized information systems of Swedish insurance companies and banks.

Against this background it appears safe to conclude that each one of the sectoral employment projections in section 4.4 is bound to clash with present regional population and employment goals, if each region roughly maintains its 1975 sectoral employment shares. The base case projections appear incompatible with the regional employment goals because it seems unlikely that enough successful policies can be organized for the outmigration of the production of public and private services from Stockholm to distant cities. All four cases also forcefully induce a more concentrated urban settlement, even if the regional balance is restored through countervailing market forces or policies.

Our conclusion is that the higher the rate of productivity growth in the tertiary sector (especially in the public services sector) compared with the manufacturing sector, and the more labor saving its technical progress, the better the possibilities are of both attaining a rapid economic growth *and* restoring a more balanced development of regional labor markets.

According to the sectoral projections, the main structural adjustment in Sweden up to the year 2000 is associated with the declining importance of the manufacturing industry compared with the tertiary sectors in particular but also with the primary sectors. Since the primary sectors and the raw material based industries have a projected slower employment decline than other manufacturing sectors, the adjustment pressures of northern Sweden merely emanate from the same problem as all Sweden compared with the metropolitan regions: the pronounced concentration of tertiary production in the Stockholm area in particular and the disruptively strong projected expansion of such production.

One feature of this projected sectoral development is the almost equiproportional contraction of all parts of the manufacturing industry. Thus according to our projections, there are no marked differences between the earlier expanding parts of the manufacturing industry and those parts that have already been contracting for some time. This feature of the projections, which is attributable to the chosen world market scenario, is in our opinion rather unrealistic. All information about the emerging changes in the international division of labor in the world market for manufacturing products suggests strong incentives to structural adjustments in industrialized countries. At present we must unfortunately accept the sectoral projections. This implies that the regional adjustment associated with these projections will be small unless both the interregional division of labor is different *within* the investigated industrial sectors *and* the growth rates differ a great deal between the subsectors at more disaggregated levels. We know from the development in the 1960s and 1970s that this is likely (Ohlsson 1979, 1980d). The projected sectoral growth pattern, however, constitutes a break with earlier sectoral trends, which makes it difficult to bring the analysis further on this point by utilizing information at more detailed subsector levels.

In conclusion, the projected sectoral changes within the manufacturing industry do not give rise to problems concerning major additional impacts from this sector on the regional balance of the domestic labor market. The world market scenarios used for the projections, however, leave much doubt about the rather optimistic outlook for raw material and raw material based production as well as raw labor intensive production compared with more technologically sophisticated products.

6 EVALUATION AND POSSIBLE ELABORATIONS OF THE METHODOLOGICAL APPROACH

The main purpose of this study is to identify possible future development paths for the Swedish economy in a context where both world market conditions and domestic factor accumulation and productivity growth are explicitly taken into account. A second purpose is to apply a slightly new approach in the analysis of these issues. Thus after the presentation of our findings concerning the substantive issues, it is appropriate to evaluate the adopted methodological approach and to point out some future directions of research.

The basic idea in our approach is to focus on the interaction between domestic and world market factors within a general equilibrium framework. This framework is represented here by a general equilibrium model of the Swedish economy. The model analysis generated two results that are suitable points of departure for an evaluation of the approach.

The first of these is the projected improvement in Sweden's terms of trade, which takes place despite a considerable projected increase in oil prices. In a technical sense our result is the combined effect of three factors: the relatively low values of the price elasticity parameters in the import and export functions, a relatively fast projected growth of world market trade, and a relatively slow domestic economic growth. That these were the key factors was confirmed by an extensive sensitivity analysis of the results.

These findings suggest that it is important to take both the supply side (world market prices) and the demand side of the rest of the world explicitly into account in the analysis. Thus the terms of trade of the Swedish economy can be determined from world market prices, in foreign currency units, only when the rates of growth in Sweden and Sweden's trading partners coincide. When this is not the case, which is the normal situation, projections of world market prices become an uncertain basis for projections of the terms of trade.

Obviously our results for the projected development of the terms of trade depend on price elasticity parameters in the import and export functions. A rather extensive sensitivity analysis, however, with relatively large variations of the import and export price elasticities around the adopted values, indicated a substantial robustness of the results with respect to these parameters. Nevertheless the treatment of foreign trade in the model might be the crucial factor behind our results. This is because the very existence of downward sloping price-dependent import and export functions can be questioned for a country like Sweden, which to a large extent conforms to the concept of a small, open economy.

In such an economy the producers in the trade-exposed sectors in general can be regarded as price takers on international markets. Available econometric evidence, however, does not generally support the small, open economy assumptions for Sweden. We will not dwell on this issue here* but only point out that both our results and the specification of the model depends on the existence of downward sloping import and export functions.

The other result that was interesting from the methodological point of view was the limited structural change within the trade-exposed sector in our projections. Thus there were only two trading sectors with a considerably different development than the trade-

*A fairly extensive discussion about this issue can be found in Bergman and Pór (forthcoming).

exposed sector as a whole. These were the shipyards and the metal industry. That is, most of the projected reallocation of resources within the trade-exposed sector can be regarded as an adjustment to comparative advantage changes that have already taken place. This points to the basic difficulty with our approach: the projections of domestic factor accumulation and productivity change might well reflect the same expectations as those underlying the projections of world market prices and trade flows. If that is the case the two sets of projections cannot be used to generate projections of future changes in comparative advantages.

Thus our limited knowledge of the expectations underlying the projections of exogenous conditions used in this study makes it difficult to draw conclusions regarding future structural change in Sweden on the basis of our results. There seems to be two ways to approach this issue. One is simply to make a closer investigation of the scenarios for domestic factor accumulation and, particularly, for productivity growth. Another is to expand the representation of the rest of the world in the model in such a way that world market prices and trade flows can be generated from explicit assumptions of production accumulation, productivity change, and demand changes in the rest of the world. These approaches are not mutually exclusive, and neither can be preferred on *a priori* grounds.

It is clear that our approach rests on the assumption that the projection of world market conditions is independent of the projection of exogenous domestic conditions. Even if this assumption is satisfied one way or another, however, the usefulness of the exercises presented in this report to a large extent depends on the properties of the model used in the analysis. Obviously the model used in this study has definite limitations. A general equilibrium model of the type used here, i.e., where factors of production can be reallocated between sectors without friction, can be used to identify the degree of structural imbalance in the economy. If, however, the equilibrium allocation of resources at one point in time differs considerably from that at another point in time, it can only be concluded that some kind of structural change process must take place if both equilibria are to be realized; the model does not say anything about the nature of this process.

Consequently, a desirable improvement of the model would be to incorporate some of the rigidities that characterize the real world. The most natural elaboration of the model in this context would be to incorporate a "putty–clay" nature of capital, thus giving the model an explicit time dimension and a specification such that sectoral reallocations of capital take place through investments.* Further elaborations could involve an explicit regional dimension and a subdivision of the labor market into a number of more or less isolated submarkets.

From our results it is obvious that the public sector plays a crucial role in industrial development projections. Little is known about the rate of productivity change of the public sector and the determinants behind this change. Perhaps even more crucial from the methodological point of view is that no policy imposed rigidities could be taken into account. Nor is the role of the government in the formation of human and non-human capital explicitly recognized in the projection model. Possible elaborations of the public sector and the role of the government appear therefore as interesting future avenues of research.

*This is done in the "dynamic" model presented in Bergman and Pór (forthcoming).

ACKNOWLEDGMENTS

The authors wish to thank Professor Peter J. Lloyd for many constructive comments on an earlier version of the report; the comments by an anonymous referee were also useful. Of course the responsibility for all remaining errors is with the authors. Financial support for part of the work from the Tercentennial Fund of the Central Bank of Sweden is gratefully acknowledged.

REFERENCES

Åberg, Y. (1969) Produktion och produktivitet i Sverige 1861–1965 (Production and Productivity in Sweden 1861–1965). Industrial Institute for Economic and Social Research. Stockholm: Almqvist and Wiksell.

Armington, P.S. (1969) A Theory of Demand for Products Distinguished by Place of Production. International Monetary Fund (IMF) Staff Papers 16:159–178.

Bergman, L. and A. Pór (1980) A Quantitative General Equilibrium Model of the Swedish Economy. WP-80-04. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Bergman, L. and A. Pór (forthcoming) Computable Models of General Equilibrium in a Small, Open Economy. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Göteborgs kommun (1978) Göteborgs kommuns Verkstadsindustriut, Näringspolitiska utredningar och förslag (The Engineering Industry Investigation of Gothenburg Commune, Industry Policy Investigations and Proposals). Gothenburg: Göteborgs kommun.

Hamilton, C. (1979) Import penetration and import price elasticities with special reference to dynamic LDC export commodities: the case of Sweden 1960–75. Seminar paper no. 116. Stockholm: Institute for International Economic Studies.

IVA and IUI (1979) Teknik och industristruktur – 70-talets ekonomiska kris i historisk belysning (Technology and industrial structure – the economic crisis of the 1970s in historical comparison). Stockholm: The Royal Academy of the Engineering Sciences (IVA) and the Industrial Institute for Economic and Social Research (IUI).

Leontief, W.W. *et al.* (1977) The Future of the World Economy: A United Nations Study. New York: United Nations.

OECD (1979) The Impact of Newly Industrializing Countries on Production and Trade in Manufacturers. Report by the Secretary General. Paris: Organization for Economic Cooperation and Development.

Ohlsson, L. (1977) Patterns of Engineering Trade Specialization 1960–1970, and Sweden's Factor Abundance. Journal of Political Economy 85(2):361–378.

Ohlsson, L. (1979) Components of Urban Industrial Employment Change in a Small, Open Economy: Sweden. WP-79-32. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Ohlsson, L. (1980a) Sweden's Specialization in Typical LDC Industries and Its Recent Industrial Intervention Policies. Mimeograph. Stockholm: The Economic Research Institute of the Stockholm School of Economics.

Ohlsson, L. (1980b) Världshandelsförändringar och Nordens framtida industrisysselsättning (World trade developments and future industrial employment in Nordic countries). Nordisk utredningsserie A(2). Stockholm: Nordiska ministerrådet.

Ohlsson, L. (1980c) Engineering Trade Specialization of Sweden and Other Industrial Countries – A Study of Trade Adjustment Mechanisms of Factor Proportions Theories. Amsterdam: North-Holland Publishing Company.

Ohlsson, L. (1980d) Industrial Structure and Possible Industrial Futures of the Malmöhus County. Mimeograph. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Restad, T. (1976) Modeller för samhällsekonomisk perspektivplanering (Models for Long-Term Indicative Economic Planning). Stockholm: Liber förlag.

DYNAMIC LINEAR PROGRAMMING MODELS OF ENERGY, RESOURCE, AND ECONOMIC-DEVELOPMENT SYSTEMS

Anatoli Propoi and Igor Zimin
International Institute for Applied Systems Analysis, Laxenburg, Austria

SUMMARY

This report develops a unified dynamic linear programming approach to studying long-range development alternatives in the energy sector. With the demand for energy and the supply of nonenergy resources needed to develop the energy supply system given exogenously, the report first seeks the optimal mix, phased over time, of different energy technologies. Next, it considers the problem of finding, for primary energy resources, the optimal mix over time of different exploration and extraction technologies. The third part of the report uses an optimization version of a dynamic input–output model to study the macroeconomic impacts of the energy sector. Finally, the report discusses the interactions among these models, presents a general dynamic linear programming framework, and takes up some related methodological issues.

INTRODUCTION

This report is an attempt to review and extend methodological research into the development of complex systems. One very typical, and probably the most urgent, example of this sort of problem is the analysis and planning of the long-term development of energy systems. During the last decade, interest in energy problems has considerably increased all over the world and we have witnessed significant progress in the field (A.A. Makarov and Melentjev 1973; Häfele and Manne 1974; Häfele 1974; Hudson and Jorgenson 1975; Häfele and Sassin 1976; Belyaev et al. 1976; Häfele and A.A. Makarov 1977; Häfele et al. 1977; A.A. Makarov 1977; Kononov 1977; Behling et al. 1977; Hoffman and Jorgenson 1977). However, most of this work has been concerned with the detailed implementation of different energy models. As regards methodological mathematical analysis of the problem, we must of course expect a slight time lag at first, but preliminary attempts have already been made (see, for example, Alta Conference 1975; Tomlin 1976).

Meanwhile, when we analyze the outputs of various energy models implemented in different ways, many methodological questions arise: for example, how should models of energy supply, resources, and the economy be linked into an overall (national) system; what is the most appropriate form of world ("global") energy model -- a game-theoretical, optimization, or simulation model; how does our uncertainty concerning future input data influence our degree of certainty about the correctness of present decisions; etc. These questions do not only relate to energy models but are also of concern for any problems involving the long-term development of a complex system (Aganbegyan et al. 1974; Aganbegyan and Valtukh 1975); one example is the analysis of the long-term interaction between manpower and economic development (Propoi 1978).

This report tries to answer some of the questions outlined above. The first three sections describe basic dynamic optimization models – of energy supply, resources, and economic development – all formally presented in a unified dynamic programming framework (Propoi 1973, 1976; Ho 1979). Section 1 considers models of Energy Supply Systems (ESS); the demand for energy and the supply of nonenergy resources needed to develop the ESS are given exogenously, and we seek the optimal mix, phased over a period of time, of different energy technologies. Section 2 examines resource models. Here the problem is to find, for primary energy resources, the optimal mix over time of different extraction and exploration technologies. Section 3 describes dynamic linear programming models of an economy; these are basically optimization versions of dynamic input–output models.

In describing these models, we have tried to concentrate on the most typical features of each, omitting the various details of implementation in order to obtain three basic formalized models which could be useful for subsequent mathematical analysis. The internal structure of the report follows on directly from this: in each of the first three sections we start by considering a basic model and then examine some related real models which can be viewed as modified versions of the basic model.

Sections 1–3 consider each model independently on a national (or regional) level. Methods for linking different models (for example, energy–economy or resources–energy) are discussed in Section 4, while Section 5 suggests a canonical form for the dynamic linear programming problem to which all the models can be reduced. The report closes with a recapitulation of the main conclusions and suggestions for further research.

This report is primarily a review, intended to give the various models a unified presentation, thus providing a basis for further development of methods for the solution and analysis of such models.

1 ENERGY SUPPLY MODELS

We begin with models of Energy Supply Systems (ESS) because ESS play central roles in any study of energy resources. The main purpose of the ESS models is to study major energy options over the next 25–50 years and longer, thus determining the optimal feasible transition from the mix of technologies for energy production currently used (basically fossil fuels), to a more progressive and, in some sense, optimal future mixture of technologies (nuclear, coal, solar, etc.) for a given region (or country). When considering

ESS models, we will basically follow the Häfele–Manne model (Häfele and Manne 1974), and then discuss different versions and modifications of the models.

In formulating Dynamic Linear Programming (DLP) problems, it is useful to identify

(i) the *state equations* of the systems with the *state* and *control* variables clearly separated
(ii) the *constraints* imposed on these variables
(iii) the *planning period* T – the *number* of periods during which the system is considered and the *length* of each period
(iv) the *performance index* (or objective function) gives some quantitative measure of the performance of a program

We will now consider these four stages separately as applied to the ESS models.

1.1 Basic Model

1.1.1 State Equations

The ESS model is broken down into two subsystems: energy production and conversion, and resource consumption. Hence, two sets of state equations are needed.

Energy production and conversion subsystem. The subsystem consists of a certain number of technologies for energy production (fossil, nuclear, solar, etc.). The state of the subsystem during each period t is described by the values of capacities during that period t for all energy-production technologies.

Let

$y_i(t)$ be the capacity of the ith energy-production technology ($i = 1, 2, \ldots, n$) during period t;
n be the total number of different technologies for energy production to be considered in the model; and
$v_i(t)$ be the increase in the capacity of the ith technology over period t ($i = 1, 2, \ldots, n$).

It is assumed that the lifetime of each unit of productive capacity, for example each power plant, is limited: this limited lifetime, characteristic of facilities based on technology i, will be denoted by τ_i.

Thus, the state equations which describe the development of the energy production and conversion subsystem will be as follows

$$y_i(t+1) = y_i(t) + v_i(t) - v_i(t - \tau_i) \qquad (i = 1, 2, \ldots, n;\ t = 0, 1, \ldots, T-1) \qquad (1.1)$$

with the given initial conditions

$$y_i(0) = y_i^0 \qquad (i = 1, 2, \ldots, n) \qquad (1.2)$$

The increase in the capacity of the ith technology, $v_i(t)$, during the period preceding the time horizon considered ($t < 0$) is also assumed to be known

$$v_i(-\tau_i) = v_i^0(-\tau_i), \ldots, v_i(-1) = v_i^0(-1) \qquad (i = 1, 2, \ldots, n) \tag{1.3}$$

where $\{v_i^0(-\tau_i), \ldots, v_i^0(-1)\}$ are given numbers.

Equations (1.1) and (1.2) can be rewritten in vector form

$$y(t+1) = y(t) + v(t) - v(t - \boldsymbol{\tau}) \tag{1.1a}$$

$$y(0) = y^0 \tag{1.2a}$$

Here

$y(t) = \{y_i(t)\}$ is a *state vector* of the subsystem in period t, describing the state of the energy production and conversion subsystem $i (i = 1, 2, \ldots, n)$ in this period;
$v(t) = \{v_i(t)\}$ is a *control vector*, describing control actions affecting subsystem $i (i = 1, 2, \ldots, n)$ in period t; and
$\boldsymbol{\tau} = \{\tau_i\}$ $(i = 1, 2, \ldots, n)$

Resource consumption subsystem. State equations of this subsystem describe the dynamics of cumulative amounts of extracted primary energy resources.

Let

$z_j(t)$ be the cumulative amount of the jth resource extracted by the beginning of period (sometimes year) t, where $(j = 1, 2, \ldots, m)$;
m be the total number of different primary resources under consideration; and
$q_{ji}(t)$ be the fraction of the jth resource (primary energy input) required for loading the capacity of the ith energy production technology (secondary energy output) in period t $(i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m)$; $q_{ji}(t)$ represents the conversion process $j \rightarrow i$.

Generally, some capacity will not always be completely loaded; therefore we introduce a new variable $u_i(t)$ which represents the degree of utilization of productive capacity based on technology i $(i = 1, 2, \ldots, n)$ in period t. It is evident that

$$u_i(t) \leqslant y_i(t) \qquad (i = 1, 2, \ldots, n) \tag{1.4}$$

or, in vectorial form

$$u(t) \leqslant y(t) \tag{1.4a}$$

If we assume that the primary energy resource extraction during period t is proportional to the degree of utilization of energy-production capacity in this period, we can write the state equations in the form

$$z_j(t+1) = z_j(t) + \sum_{i=1}^{n} q_{ji}(t)\, u_i(t) \tag{1.5}$$

with initial conditions

$$z_j(0) = z_j^0 \qquad (j = 1, 2, \ldots, m) \tag{1.6}$$

or, in matrix form

$$z(t+1) = z(t) + \mathbf{Q}(t)u(t) \tag{1.5a}$$

$$z(0) = z^0 \tag{1.6a}$$

Here $z(t)$ is a state vector and $u(t)$ is a control vector. The subsystems (1.1) and (1.5) are linked by means of the inequalities (1.4).

If the conversion process $j \to i$ is denoted by the matrix $\tilde{\mathbf{Q}}(t) = \{\tilde{q}_{ij}(t)\}$, then eqn. (1.5a) should be rewritten as

$$z(t+1) = z(t) + \tilde{\mathbf{Q}}^{\mathrm{T}}(t)u(t) \tag{1.5b}$$

where $\tilde{\mathbf{Q}}^{\mathrm{T}}$ denotes the transpose of the matrix $\tilde{\mathbf{Q}}$.

In some cases it is necessary to introduce variables representing stocks of the primary resources extracted (inventory resources). Let $\tilde{z}_j(t)$ be such a variable for the jth resource and $w_j(t)$ the amount of this resource extracted annually. The state equation for the inventory subsystem will then be as follows

$$\tilde{z}(t+1) = \tilde{z}(t) + w(t) - \mathbf{Q}(t)u(t)$$

In the above case, $\tilde{z}(t) = 0$ for all t, and $w(t) = \mathbf{Q}(t)u(t)$. This is a reasonable assumption because, in the long term, one can neglect the accumulation of stocks of resources.

It should be noted that the real equations of the resource-consumption subsystem are more complex [see Häfele and Manne (1974) and the discussion in Section 1.2].

1.1.2 Constraints

The state equations (1.1) and (1.5) specify the dynamic constraints on variables, but we also have a number of static constraints on variables for each period t.

Nonnegativity. It is evident that no variables introduced into the state equations (1.1) and (1.5) can be negative

$$v(t) \geqslant 0,\ y(t) \geqslant 0,\ u(t) \geqslant 0,\ z(t) \geqslant 0 \tag{1.7}$$

Availability. To begin with, upper limits are imposed on the annual construction rates

$$v_i(t) \leqslant \bar{v}_i(t) \qquad (i = 1, 2, \ldots, n) \tag{1.8}$$

where $\bar{v}_i(t)$ are given numbers. In a more general form, these constraints may be written as

$$\mathbf{F}(t)v(t) \leqslant f(t) \tag{1.9}$$

where $f(t)$ is the vector of nonenergy inputs which are needed for the energy production subsystem. The matrix $\mathbf{F}(t)$ denotes the amounts of these resources required for the construction of one unit of capacity using the ith technology in period t. Limits on the rates of introduction of new technology can also be written in the form of eqns. (1.8) or (1.9). More general cases, where the time lags between investment decisions and actual increases in capacity are taken into account, are considered in Section 3.1. In such a situation we can directly link the ESS model with the economic model described in Section 3.

The constraints on the availability of the primary energy resources may be given in the form

$$z(t) \leqslant \bar{z}(t) \tag{1.10}$$

where $\bar{z}(t)$ is the vector of all available energy resources (resources in place) in period t, and $z(t)$ is calculated from eqn. (1.5).

The constraints on the availability of the secondary energy-production capacities are given by inequality (1.4).

Demand. The intermediate and final demands for energy are assumed to be given for all planning periods considered. Hence the demand constraints can be written as

$$\sum_{i=1}^{n} d_{ki}(t)u_i(t) \geqslant d_k(t) \tag{1.11}$$

or

$$\mathbf{D}(t)u(t) \geqslant d(t) \tag{1.11a}$$

where

$d(t) = \{d_k(t)\}$ is the given vector for all t ($t = 0, 1, \ldots, T-1$) of energy demand, both intermediate and final (that is, including both the electrical and nonelectrical components of final demand); and

$\mathbf{D}(t) = \{d_{ki}(t)\}$ is the matrix with components $d_{ki}(t)$, defining either intermediate consumption of secondary energy k per unit of total secondary-energy production, or the conversion efficiency when producing one unit of secondary energy k from energy originally produced using technology i.

1.1.3 Planning Period

The planning period is broken down into T steps, where T is given exogeneously. Each step is of a certain length (e.g., one, three, or five years). Häfele and Manne (1974) chose a planning period of 75 years and each step corresponded to three years, so in that case $T = 25$. Since information on the coefficients of the model becomes more inaccurate with an increasing number of steps it is useful to consider steps which are not all of equal length. For example, Marcuse et al. (1976) decided on a planning period of 100 years, divided into ten steps of varying length (the first five periods of six years each, the next three periods of ten years each, and the last two periods of twenty years each).

1.1.4 Objective Function

The choice of the objective function is one of the more important stages in model building. Full discussion of the economic aspects of ESS modeling objectives is beyond the scope of this report. Here we would like specifically to emphasize only two points: first, in many cases the objective functions can be expressed as linear functions of state and control variables, thus making it possible to use Linear Programming (LP) techniques. Second, the optimization procedure should not be viewed as a final part of the planning process (yielding a "unique" optimal solution), but only as a tool for analyzing the connection between policy alternatives and system performance. Thus in practical applications a policy analysis with various different objective functions is required. For our purpose, however, it is sufficient to limit ourselves to some typical examples of objectives.

Let us consider the objective function which expresses the total capital costs, discounted over time, for both the construction and the operation of units of productive capacity based on technology i

$$J = \sum_{t=0}^{T-1} \beta(t) \left[\sum_{i=1}^{n} c_i^u(t) u_i(t) + \sum_{i=1}^{n} c_i^v(t) v_i(t) \right] \tag{1.12}$$

where

- $c_i^u(t)$ are the operating and maintenance costs for units of productive capacity based on technology i in period t;
- $c_i^v(t)$ are the investment costs for units of productive capacity based on technology i in period t; and
- $\beta(t)$ is the discount rate.

We can express this in vector form as

$$J = \sum_{t=0}^{T-1} \beta(t) \; [(c^u(t), u(t) \; + (c^v(t), v(t)] \tag{1.12a}$$

Note that the scalar product $(c^u(t), u(t))$ expresses not only direct operating and maintenance costs during step t but may also indirectly include the cost of primary resources consumed during this step. In a more explicit way, this cost can be written as $(c^u(t), Q(t)u(t))$, where $c^u(t)$ should increase with the cumulative amount of resources consumed. This leads to a nonlinear objective function. A reasonable approximation in this case is a stepwise function for $c^u(t)$. Thus, $c^u(t)$ in eqn. (1.12) can be a stepwise function, with values for each step which depend on the values of cumulative extraction of resources $z(t)$ [or on the difference $\bar{z}(t) - z(t)$].

1.1.5 Statement of the Problem

To begin with, we introduce a number of definitions. A sequence of vectors

$$v = \{v(0), \ldots, v(T-1)\}, \; u = \{u(0), \ldots, u(T-1)\}$$

are *controls* of the system; a sequence of vectors

$$y = \{y(0), \ldots, y(T)\}$$

determined by eqns. (1.1) and (1.2) defines a (*capacity*) *trajectory* of the system; and a sequence of vectors

$$z = \{z(0), \ldots, z(T)\}$$

determined by eqns. (1.5) and (1.6) is a (*cumulative resources*) *trajectory* of the system.

Sequences of control and state vectors $\{v, u, y, z\}$ which satisfy all the constraints of the problem [for example eqns. (1.1)–(1.11) in this case] are called *feasible*. Having chosen *feasible controls* v and u one can obtain, by using eqns. (1.1)–(1.3), (1.5), and (1.6), *feasible state trajectories* y and z. Thus

$$J = J(y(0), z(0), v, u) = J(v, u)$$

A feasible control $\{v^*, u^*\}$ which minimizes the objective function described in eqn. (1.12) or the equation above will be called an *optimal control*.

We can now formulate the optimization problem for the energy supply system.

Problem 1.1. Given the state equations

$$y(t+1) = y(t) + v(t) - v(t-\tau) \tag{1.1a}$$

$$z(t+1) = z(t) + \mathbf{Q}(t)u(t) \tag{1.5a}$$

with initial conditions

$$y(0) = y^0 \tag{1.2a}$$

$$z(0) = z^0 \tag{1.6a}$$

and known parameters

$$v(-\tau) = v^0(-\tau), \ldots, v(-1) = v^0(-1) \tag{1.3}$$

find controls $\{v, u\}$, and corresponding trajectories $\{y, z\}$, which satisfy the constraints $v(t) \geqslant 0$; $u(t) \geqslant 0$; $y(t) \geqslant 0$; and $z(t) \geqslant 0$

$$\mathbf{F}(t)v(t) \leqslant f(t) \tag{1.9}$$

$$u(t) \leqslant y(t) \tag{1.4a}$$

$$z(t) \leqslant \bar{z}(t) \tag{1.10}$$

$$\mathbf{D}(t)u(t) \geqslant d(t) \tag{1.11a}$$

and minimize the objective function

$$J(v, u) = \sum_{t=0}^{T-1} \beta(t)[(c^u(t), u(t)) + (c^v(t), v(t))] \tag{1.12a}$$

Verbally, the policy analysis in the energy supply system model, which is formalized as Problem 1.1, can be stated as follows.

At the beginning of the planning period, energy production capacities broken down into several "homogeneous" technologies (fossil, nuclear, solar, etc.) are known [eqn. (1.2a)]. There are various possible options for developing these initial energy production capacities in the system during the period considered. These options are subject to constraints on the availability of primary energy resources [eqns. (1.5a), (1.6a), (1.10)], and constraints on the availability of nonenergy resources [eqn. (1.9)] required for the construction of new units of energy production capacity. Each of these options has its own advantages and disadvantages, and the problem consists of finding an optimal mix of these options, which, over a given period,

- meets the given demand for secondary energy [eqn. (1.11a)]
- satisfies the constraints on the availability of primary energy resources and nonenergy resources [eqns. (1.9), (1.10)]
- minimizes the total costs (for both construction and operation) [eqn. (1.12a)]

There are two important vector parameters in the model, both of which are given exogenously: the amount of nonenergy resources $f(t)$ available during the planning period, and the demand $d(t)$ for secondary energy. These values mainly affect the interaction of the energy supply system with the economic development system (see Section 4).

1.2 Discussion

The version of an energy supply system (ESS) model considered above is somewhat simplified, but nevertheless it reveals the major features of real systems. The actual implementation of the various ESS models is naturally more detailed and complicated; it depends to a great extent on the general approach selected for the overall ESS model, and on the assumptions about energy and the economy used for building its separate submodels. We will not, however, pay too much attention to the physical peculiarities of different ESS models but will rather try to emphasize the methodological characteristics of the various models and their relationships to Problem 1.1. It should be noted that some of the notation used below is different to that used in the original versions of the models to facilitate analysis and comparison.

1.2.1 Häfele–Manne Model

To illustrate the model described above, we will consider the Häfele–Manne model (Häfele and Manne 1974; Suzuki 1975) in rather more detail. In the model a 75-year planning horizon is subdivided into 25 intervals, each three years in length. Total energy production capacity is divided into two groups: new technologies, for which additional capacity is being constructed during the planning horizon and some "old" technologies. We denote the vectors of new and old capacities by $y(t) = \{y_i(t)\}$ $(i = 1, 2, \dots, n)$ and $y_0(t) = \{y_{0i}(t)\}$ $(i = 1, 2, \dots, n_0)$, respectively. The vector $y(t)$ refers to capacity installed or added to during the planning horizon and based on such technologies as coal (COAL), petroleum, gas, etc. (PETG), the light water reactor (LWR), the fast breeder reactor (FBR), electrolytic production of hydrogen (ELHY), etc.; the exogenous vector $y_0(t)$ refers to the amount of capacity based on fossil fuels (coal, petroleum, gas, etc.) already available at the beginning of the planning horizon. It is assumed that all units of new capacity are retired after 30 years of service, and that they are operated at a constant rate throughout the 30-year period. Thus, the state equations for the energy production subsystem can be written in the form of eqn. (1.1), where i = COAL, PETG, LWR, FBR, ELHY, etc; $t = 0, 1, \ldots, 24$; $\tau_i = 10$ for all i; and $v_i(t)$ is the increase in the capacity of the ith technology in the three years included in time period t [by assumption $v_i(t) = 3\tilde{v}_i(\tilde{t})$, where $\tilde{v}_i(\tilde{t})$ is the annual increase in year $\tilde{t} = 0, 3, 6, \ldots$].

Häfele and Manne (1974) assume a total loading of capacities

$$u_i(t) = y_i(t) \tag{1.13}$$

In this case, the state equations for the energy consumption subsystem have the form

$$z_j(t+1) = z_j(t) + a_j[y_j(t) + y_{0j}(t)] \qquad (j = \text{COAL}, \text{PETG}) \tag{1.14}$$

for coal and for petroleum and gas; in other words, the cumulative consumption $z_j(t+1)$ of coal or petroleum and gas by the beginning of period $t + 1$ is equal to the cumulative consumption $z_j(t)$ of this resource by the beginning of period t plus the consumption by the existing production capacity $y_j(t) + y_{0j}(t)$ during period t.

For natural uranium (NU) we have the equation

$$\begin{aligned} z_{\text{NU}}(t+1) = z_{\text{NU}}(t) &+ [a^1_{\text{NU}} y_{\text{LWR}}(t) - a^2_{\text{NU}} y_{\text{NUHC}}(t)] \\ &+ b^1_{\text{NU}}[v_{\text{LWR}}(t+1) - v_{\text{LWR}}(t-10)] \\ &+ b^2_{\text{NU}}[v_{\text{HTR}}(t+1) - v_{\text{HTR}}(t-10)] \end{aligned} \tag{1.15a}$$

Examining the terms on the right-hand side of eqn. (1.15a), we see first that natural uranium is required in period t for the current refueling of existing light water reactor (LWR) capacity; we note also that part of the total requirement can be met by using *high cost* natural uranium (NUHC), which therefore appears as a negative term. Additional amounts of natural uranium are required for setting up new LWR and HTR (high temperature reactor) capacity three years later (in the next period, $t + 1$); because the spent fuel is reprocessed, uranium is effectively released when the LWR and HTR facilities are retired at the end of their service lifetime of ten three-year periods [this accounts for the negative terms $v_{\text{LWR}}(t-10)$ and $v_{\text{HTR}}(t-10)$, respectively].

For natural uranium it is appropriate to speak of cumulative resource *consumption*, but for man-made plutonium we must consider cumulative resource *production*, which alters the sense of the state equation. For plutonium the state equation includes the following elements. The cumulative sum $[z_{\text{PLUT}}(t+1)]$ of plutonium produced by the beginning of period $t + 1$ is equal to the cumulative sum $[z_{\text{PLUT}}(t)]$ of the plutonium produced by the beginning of period t, plus production $[y_{\text{LWR}}(t)]$ from LWRs during period t, plus the gain $[y_{\text{FBPL}}(t)]$ from fast breeder reactors (FBRs) during period t, plus amounts

$[\nu_{FBR}(t-10)]$ "reclaimed" from FBRs retired at the end of their 30-year lifespan, minus consumption $[\nu_{FBR}(t)]$ for setting up new FBR capacity during period t. Stating this mathematically

$$z_{PLUT}(t+1) = z_{PLUT}(t) + a^1_{PLUT}y_{LWR}(t) + a^2_{PLUT}y_{FBLP}(t) + b_{PLUT}[\nu_{FBR}(t-10) - \nu_{FBR}(t)] \quad (1.15b)$$

It should be emphasized that these equations are given here only for illustration: complete explanation of the equations would require a description of the nuclear cycle, which would fall outside the scope of this report [for further details see, for example, Häfele and Manne (1974)]. Here we will merely state that in matrix form these equations may be written as

$$\mathbf{Z}(t+1) = \mathbf{Z}(t) + \mathbf{A}_0 y_0(t) + \mathbf{A}_1 y(t) + \mathbf{A}_2 y(t-1) + \mathbf{B}_1 \nu(t) + \mathbf{B}_2 \nu(t+1) - \mathbf{B}_3 \nu(t-\tau) \quad (1.15c)$$

and over the long term they can in fact be reduced to eqn. (1.5).

Demand constraints in the model (Häfele and Manne 1974) are written in the form

$$\mathbf{D}y(t) + \mathbf{D}_0 y_0(t) \geqslant d(t) \quad (1.16)$$

for final demand and

$$\mathbf{D}_1 y(t) + \mathbf{D}_2 [\nu(t+1) - \nu(t-\tau)] \geqslant 0 \quad (1.17)$$

for intermediate demand. Only two types of demand are considered, namely, for electrical and nonelectrical energy. Häfele and Manne give the objective function in a linear form similar to eqn. (1.12) for their model societies 1 and 2, and in a nonlinear form

$$J = \sum_{t=0}^{T-1} [a_1(t) d_1^{b_1}(t) + a_2(t) d_2^{b_2}(t)] \quad (1.18)$$

for their model society 3. In the last case it is assumed that demands [$d_1(t)$ for electrical and $d_2(t)$ for nonelectrical energy] are responsive to prices and hence are endogenously determined in the model.

1.2.2 ETA Model

The model for Energy Technology Assessment (ETA) is closely related to the energy supply system model considered above. The model was developed by Manne (1976, 1977) and represents a further development of the nonlinear version (model society 3) of the Häfele–Manne model. ETA is a medium-sized, nonlinear programming model (with linear constraints). It contains, for a 15-stage planning horizon (each stage 5 years long), a total of 300 rows, 700 columns, and 2500 nonzero matrix elements. The model was solved using MINOS – a general-purpose production code developed by Murtagh and Saunders (1978) for solving large-scale nonlinear programs with linear constraints; the code is based on the reduced-gradient algorithm and, on an IBM 370/168, takes 70 seconds to solve the first case and 30 seconds for each subsequent case (Manne 1976, 1977).

Formally, the ETA model constraints have the form of eqns. (1.1)–(1.3) and (1.13–(1.17). The objective function may be viewed in either of two equivalent ways: maximizing the sum of consumers' plus producers' surplus, or minimizing the sum of the costs of conservation measures plus interfuel substitution costs plus the costs of energy supply. In the latter case it is essentially a combination of eqns. (1.12) and (1.18). Because the objective function is formulated in this way, ETA automatically allows for price-induced conservation and also for interfuel substitution.

1.2.3 MESSAGE

The models considered above (Problem 1.1) are formally DLP models of general type (one-index models). By introducing energy flows (from supply points to demand points) we arrive at DLP models of the transportation type (two-index models). The energy models MESSAGE (Agnew et al. 1978a, b) and DESOM [see Marcuse et al. (1976) and Section 1.2.4 below] can both be written in this form. It should be noted that such models cannot be directly handled by transportation or network algorithms, and that therefore conventional LP-packages were used for their solution (Agnew et al. 1978b; Marcuse et al. 1976). The extension of transportation algorithms to handle this particular type of problem was reported recently by Krivonozhko and Propoi (1979).

MESSAGE (Model for Energy Supply Systems Alternatives and their General Environmental impact) was developed by Voss, Agnew, and Schrattenholzer at the International Institute for Applied Systems Analysis (IIASA) as an extension of the Häfele–Manne model. The model differs from its predecessors (Häfele and Manne 1974; Suzuki 1975) in that it includes all allocated secondary energy to end users, incorporates an increased number of supply technologies, makes distinctions between different price categories of natural resources, and adds the costs of resources extracted to the objective function (Agnew et al. 1978a, b).

A simplified diagram of the MESSAGE model is presented in Figure 1. Each conversion process is linked to the other blocks of the system by flows of energy inputs and outputs. Each primary energy resource is either converted into a secondary energy form by a central-station conversion process (e.g., coal converted to electricity) or used directly as a fuel by a decentralized conversion process or end-use technology (e.g., coal used for space heating).

We will now describe a very simplified version of the energy flow model.

Let $x_{jil}(t)$ be the energy flow in period t from supply category j (e.g., primary resource j) to demand category l (e.g., end-use technology l) using conversion process i. Then, following the usual procedure for transportation problems, we can define the supply of energy l which should be greater than or equal to the given demand $d_l(t)$

$$\sum_{j,i} \alpha_{jil} x_{jil}(t) \geqslant d_l(t) \tag{1.19}$$

On the other hand, the total consumption $w_j(t)$ of primary energy resource j in period t is limited by the availability of this resource

$$\sum_{i,l} \beta_{jil} x_{jil}(t) = w_j(t) \tag{1.20}$$

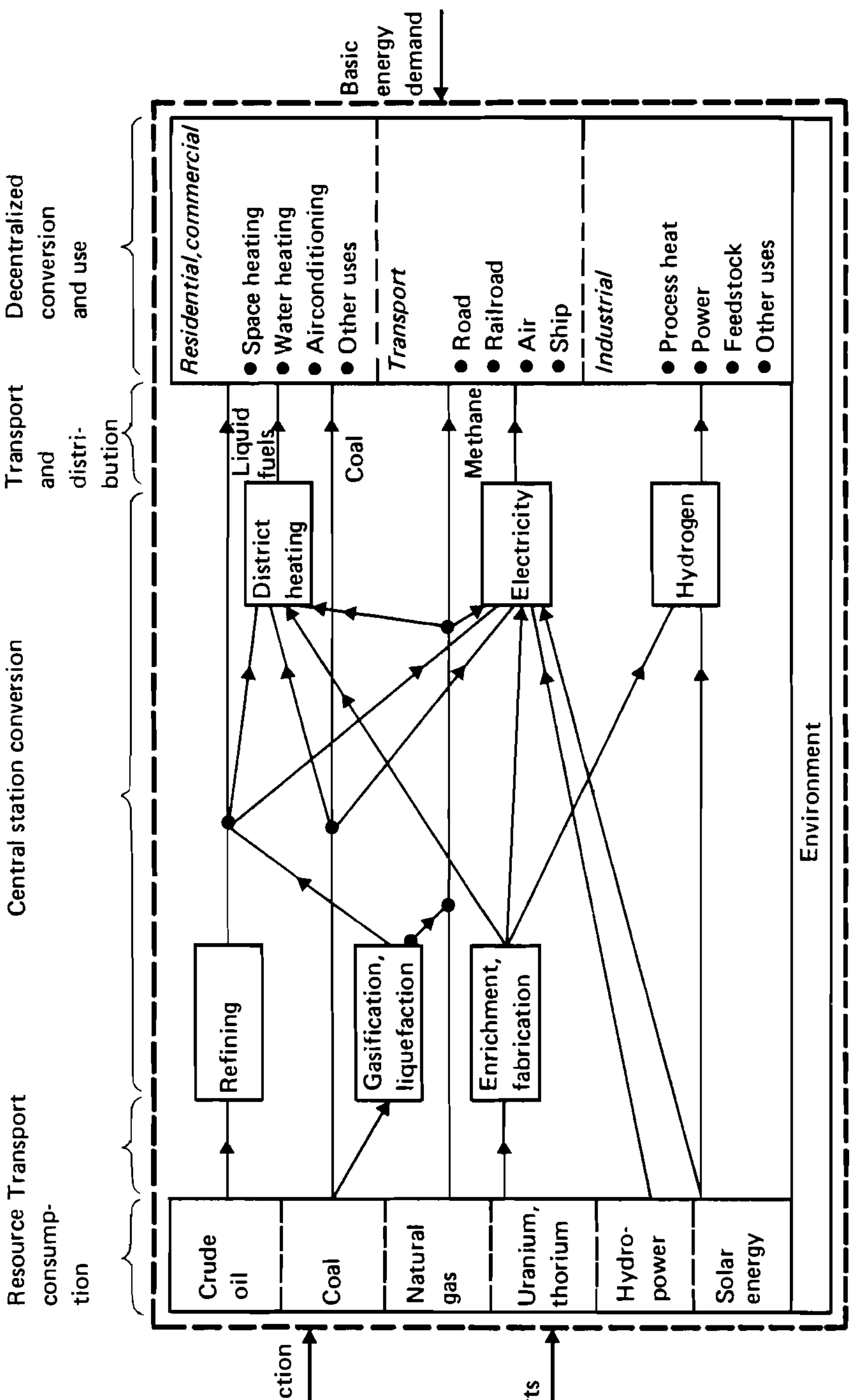


FIGURE 1 Simplified structure of the model MESSAGE.

$$z_j(t+1) = z_j(t) + w_j(t) \tag{1.21}$$

$$z_j(t) \leqslant \bar{z}_j(t) \tag{1.22}$$

Here $z_j(t)$ and $\bar{z}_j(t)$ have the same meaning as in eqns. (1.5) and (1.10). The degree of utilization $u_i(t)$ of process i is also limited by the available production capacity $y_i(t)$

$$\sum_{j,l} \gamma_{jil} x_{jil}(t) = u_i(t) \tag{1.23}$$

$$u_i(t) \leqslant y_i(t) \tag{1.24}$$

In eqns. (1.19), (1.20), and (1.23) α_{jil}, β_{jil}, and γ_{jil} are coefficients of energy-resource conversion efficiency (for examples see the next sections).

The development of the production capacity subsystem is described by state equations similar to eqn. (1.1).

We are now in a position to formulate a DLP model as follows.

Problem 1.2. Given the state equations

$$y_i(t+1) = y_i(t) + v_i(t) - v_i(t - \tau_i) \qquad (i = 1, 2, \ldots, n)$$

$$z_j(t+1) = z_j(t) + w_j(t) \qquad (j = 1, 2, \ldots, m)$$

with the initial conditions

$$y_i(0) = y_i^0;\ z_j(0) = z_j^0;\ v_i(t - \tau_i) = v_i^0(t - \tau_i)$$

$$(i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m;\ t < \tau_i)$$

find controls $\{x_{jil}(t)\}, \{v_i(t)\}$ and corresponding state variables $\{y_i(t)\}$, $\{z_j(t)\}$ which satisfy the conditions

$$u_i(t) = \sum_{j,l} \gamma_{jil} x_{jil}(t) \leqslant y_i(t)$$

$$w_j(t) = \sum_{i,l} \beta_{jil}(t) x_{jil}(t)$$

$$d_i(t) \leqslant \sum_{j,l} \alpha_{jil} x_{jil}(t)$$

$$v_i(t) \leqslant \bar{v}_i(t);\ z_j(t) \leqslant \bar{z}_j(t)$$

$$v_i(t) \geqslant 0,\ x_{jil}(t) \geqslant 0;\ y_i(t) \geqslant 0;\ z_j(t) \geqslant 0$$

and minimize the objective function

$$J = \sum_{t=0}^{T-1} \beta(t) \left[\sum_{i=1}^{n} c_i^1 u_i(t) + \sum_{i=1}^{n} c_i^2 v_i(t) + \sum_{j=1}^{m} c_j^3 w_j(t) \right] \tag{1.25}$$

The typical dimensions of the MESSAGE model are as follows. The planning horizon T is 65 years, divided into 13 periods of five years each. The numbers of each type of constraint are: demand, $7 \times T$; resources, $5 \times T$; total availability of resources, 17×1; intensity of resource extraction, $2 \times T$; and capacity loading, $5 \times T$; in addition, there are $35 \times T$ equations for capital stocks. Together with the other constraints this gives us, in terms of conventional LP problems, about 1100 rows and 1200 columns, with some 90 constraints for each period.

1.2.4 DESOM

DESOM (Dynamic Energy System Optimization Model) (Marcuse et al. 1976) was developed at the Brookhaven National Laboratory and is an extension of the Brookhaven Energy System Optimization Model (BESOM) which was a static, single-period LP model. In DESOM the demand sector has been disaggregated into technology-related end uses (22 mutually-exclusive end uses as defined by their energy-conversion processes). The general structure of DESOM is similar to that outlined in Problem 1.2.

Let us consider the state equations for the development of capacity of type i in the form

$$y_i(t+1) = y_i(t) + v_i(t) - v_i(t - \tau_i) - v_{0i}(t) \tag{1.26}$$

where the meaning of the control $v_i(t)$ and state $y_i(t)$ variables is the same as in eqn. (1.1); $v_{0i}(t)$ is the exogenously given decrease in existing (old) capacity of type i during period t.

Marcuse et al. (1976) introduced a scenario variable $\alpha(t)$ which restricts the rate of growth of capacity

$$y_i(t+1) \leqslant \alpha(t) y_i(t) \tag{1.27}$$

Generally $\alpha(t)$ is greater than one, which implies that installed capacity may expand during period t; if $\alpha(t)$ is less than one then capacity will decrease during period t.

Using eqn. (1.26) one can rewrite inequality (1.27) in the following form, which is similar to the inequality given by Marcuse et al.

$$y_{0i}(t+1) + \sum_{g=t-\tau_i}^{t} v_i(g) \leqslant \alpha(t) \left[y_{0i}(t) + \sum_{g=t-1-\tau_i}^{t-1} v_i(g) \right]$$

where

$$y_{0i}(t) = y_i(0) - \sum_{g=0}^{t-1} v_{0i}(g)$$

is the inherited capacity (capital stock of old capacities) for conversion process i at the beginning of period t (given exogenously).

To link the production subsystem with the resource-consumption subsystem, Marcuse et al. introduced demand and other constraints on intermediate energy flows. Each intermediate energy flow has associated with it a demand efficiency and a supply efficiency. The demand efficiency measures the energy loss as the intermediate flow is converted into a final energy product; the supply efficiency measures the energy loss from the point of extraction of the primary energy source to the intermediate energy flow.

If we let $x_{kl}(t)$ be the amount of intermediate energy flow in period t from supply category k to meet final energy demand l, we can define

$$u_i(t) = (1/\Delta) \sum_{(k,l)\in\Omega(i)} [x_{kl}(t)/r_{kl}] \tag{1.28}$$

where

- r_{kl} is the load factor for intermediate energy flow from supply category k to final demand category l;
- Δ is the length of period, generally, $\Delta = \Delta(t)$;
- $\Omega(i)$ is the set of indices (k, l), which defines the path of intermediate energy flow from supply k to final demand l associated with conversion process i; and
- $u_i(t)$ is the amount of installed capacity for conversion process i required in period t to deliver $x_{kl}(t)$, in other words, $u_i(t)$ is the degree of utilization of conversion process i in period t.

Evidently, the amount of installed capacity available in period t must be sufficient to produce intermediate energy flows which utilize the capacity for conversion process i in period t

$$(1/\Delta) \sum_{(k,l)\in\Omega(i)} [x_{kl}(t)/r_{kl}] \leqslant y_i(t) \tag{1.29}$$

which is similar in form to inequality (1.4).

Capacity is required to meet both base-load and peak demands in the electrical sectors. Off-peak electrical intermediate energy flows that use capacity installed for peak requirements are not included in inequality (1.29). For electricity-conversion processes

$$(1/\Delta) \sum_{(k,l)\in\Omega(i)} x_{kl}(t) \leqslant q_i y_i(t) \tag{1.30}$$

where q_i is an overall load factor, applied to all electrical capacity, which states that a conversion facility of type i can only operate for a proportion q_i of the time.

By introducing intermediate energy-flow variables it is possible to write down demand and resource constraints. The total amount of energy from intermediate energy flows $x_{kl}(t)$ must be sufficient to meet the demands $d_l(t)$

$$\sum_k d_{kl} x_{kl}(t) = d_l(t)$$

for each demand category l. Here the d_{kl} are demand coefficients representing the overall technical efficiency of a conversion technology for some intermediate energy flow from supply category k to meet final energy demand l.

On the other hand, intermediate energy flows $x_{kl}(t)$ in period t define a demand for primary energy resource j

$$\sum_{k,l} s_{jkl} x_{kl}(t) = w_j(t) \tag{1.31}$$

where

- s_{jkl} are supply coefficients representing the overall technical efficiency of the conversion technology for intermediate energy flow based on resource j from supply k to final demand l; and
- $w_j(t)$ is the amount of resource j used in period t.

Introducing the cumulative amount $z_j(t)$ of resource j extracted by the beginning of period t, one can write the state equation for the resource-consumption subsystem in the form

$$z_j(t+1) = z_j(t) + w_j(t);\ z_j(0) = z_j^0 \tag{1.32}$$

which is similar in form to eqn. (1.5). It is also evident that

$$z_j(t+1) = z_j(0) + \sum_{g=0}^{t} w_j(g)$$

Marcuse et al. (1976) built into DESOM upper and lower limits on cumulative resource extraction

$$\underline{z}_j \leqslant z_j(t) \leqslant \bar{z}_j \tag{1.33}$$

$\bar{z}_j$ is associated with the real world availability of resource j, whereas the lower limit $\underline{z}_j$ assures some minimum consumption. In addition to the constraints (1.33), DESOM contains a restriction on the rate of growth of resource extraction, namely that the amount of resource j extracted in period $t+1$ must be no greater than $\beta_j(t)$ times the amount of resource j extracted in period t

$$w_j(t+1) \leqslant \beta_j(t) w_j(t) \tag{1.34}$$

Generally $\beta_j(t) > 1$; to simulate the phasing out of a resource over time one can set $\beta_j(t) \leqslant 1$ for later periods.

As in other models, DESOM contains environmental constraints, which are written in the form

$$\sum_{k,l} e_{klm} x_{kl}(t) \leqslant E_m(t) \tag{1.35}$$

where

- e_{klm} is the amount of emission of type m for intermediate energy flow from k to l; and
- $E_m(t)$ is the maximum permissible amount of emission of type m in period t.

The objective of the problem is to minimize the total discounted cost, i.e.

$$J = \sum_{t=0}^{T-1} \gamma(t) \left[\sum_{kl} c_{kl}^1(t) x_{kl}(t) + \sum_i c_i^2(t) v_i(t) + \sum_j c_j^3(t) w_j(t) \right] \tag{1.36}$$

where

$c^1_{kl}(t)$ is the cost for intermediate energy flows (undiscounted);
$c^2_i(t)$ is the annual cost during period t for building capacity for conversion process i; and
$c^3_j(t)$ is the cost for resource j in period t.

Consideration of the variables $v_i(t)$ in the last time period is in fact incorporated in DESOM but is not shown in eqn. (1.36).

Thus the optimization problem for the DESOM model can be formulated as follows.

Problem 1.3. Given the state equations

$$y_i(t+1) = y_i(t) + v_i(t) - v_i(t-\tau_i) - v_{0i}(t)$$

$$z_j(t+1) = z_j(t) + w_j(t)$$

with initial states

$$y_i(0) = y_i^0$$

$$z_j(0) = z_j^0$$

and known parameters

$$v(-\tau_i) = v^0(-\tau_i), \ldots, v(-1) = v^0(-1)$$

$$v_{0i}(t) \qquad (t = 0, 1, \ldots, T-1)$$

find controls $\{v_i(t)\}$, $\{w_j(t)\}$, $\{x_{kl}(t)\}$, and corresponding trajectories $\{y_i(t)\}$, $\{z_j(t)\}$, which satisfy the constraints

$$v_i(t) \geqslant 0; \; x_{kl}(t) \geqslant 0; \; y_i(t) \geqslant 0; \; z_j(t) \geqslant 0$$

$$\sum_k d_{kl} x_{kl}(t) = d_l(t)$$

$$\sum_{k,l} s_{jkl} x_{kl}(t) = w_j(t)$$

$$(1/\Delta) \sum_{k,l} [x_{kl}(t)/r_{kl}] \leqslant y_i(t) \qquad (i \in I^1)$$

$$(1/\Delta) \sum_{k,l} x_{kl}(t) \leqslant q_i y_i(t) \qquad (i \in I^2)$$

$$\underline{z}_j(t) \leqslant z_j(t) \leqslant \bar{z}_j(t)$$

$$y_i(t+1) \leqslant \alpha(t) y_i(t)$$

$$w_j(t+1) \leqslant \beta_j(t) w_j(t)$$

and minimize the objective function

$$J = \sum_{t=0}^{T-1} \gamma(t) \left[\sum_{kl} c^1_{kl}(t) x_{kl}(t) + \sum_i c^2_i(t) v_i(t) + \sum_j c^3_j(t) w_j(t) \right]$$

On examination of Problem 1.3, one can see that it is very similar to those considered earlier [if we exclude the special method of introducing the intermediate flows $x_{kl}(t)$].

As reported by Marcuse et al. (1976), the model without environmental constraints had 130 row constraints and 750 variables per period. The first version of the model contains a four-period optimization problem and it takes about 30 minutes to solve on an IBM 370/155. A standard base case is being developed; this case will cover the 100-year period from 1973 to 2073. It will consist of six five-year periods to provide considerable detail from now until the turn of the century; three ten-year periods to allow for the simulation of large-scale introduction of fusion and solar technologies in the early 21st century, and finally two twenty-year periods to reduce truncation effects.

A new version of DESOM, the MARKet ALlocation Model (MARKAL), has been developed recently at the Brookhaven National Laboratory (Kydes 1978). MARKAL is currently being used by the International Energy Agency in planning strategic energy options.

1.2.5 SPI Model

This model has been developed (A.A. Makarov and Melentjev 1973; Belyaev et al. 1976; A.A. Makarov 1977; Kononov 1977; Häfele and A.A. Makarov 1977) at the Siberian Power Institute (SPI), Siberian Department of the USSR Academy of Sciences, to analyze possible energy development strategies and to compare the trends in different branches of science and technology. The model is part of a system of models for long-term energy development forecasting (for a time horizon of 30–40 years). As this system of models has already been described at length elsewhere, we will discuss here only the more important features of the SPI energy supply systems model.

The SPI model has a specific block structure with detailed descriptions, for each region k and year t, of the production, interconnection, and conversion of energy at all stages ranging from the extraction of primary energy (different kinds of fossil fuel, nuclear fuel, hydro, solar, geothermal energy), via the production and distribution of secondary energy (liquid, solid, and gaseous fuels, secondary nuclear fuel, electrical energy, steam, hot water), to the production of final energy utilized in industry, transport, agriculture, and the municipal and service sectors. For each year t the model consists of oil, coal, gas, nuclear, and electrical energy blocks; for each region k it consists of fuel and electrical energy supply blocks. Each block can be generated, introduced into a computer, and updated independently.

For each region k and year t the balance equations for production and distribution are as follows.

For primary energy α

$$\sum_{j(\alpha)} a^k_{\alpha j}(t) x^k_{\alpha j}(t) + \sum_{k'} x^{kk'}_{\alpha}(t) = \sum_{j(\beta)} b^k_{\alpha j}(t) x^k_{\beta j}(t) + \sum_{k'} b^{kk'}_{\alpha}(t) x^{kk'}_{\alpha}(t) + d^k_{\alpha}(t)$$

For secondary energy β

$$\sum_{j(\beta)} a_{\beta j}^k(t) x_{\beta j}^k(t) + \sum_{k'} x_\beta^{kk'}(t) = \sum_{j(\gamma)} b_{\beta j}^k(t) x_{\gamma j}^k(t) + \sum_{k'} b_\beta^{kk'}(t) x_\beta^{kk'}(t) + d_\beta^k(t)$$

For final energy γ

$$\sum_{j(\gamma)} a_{\gamma j}^k(t) x_{\gamma j}^k(t) = d_\gamma^k(t)$$

The various terms in the balance equations have the following meanings.

$x_{\alpha j}^k(t), x_{\beta j}^k(t), x_{\gamma j}^k(t)$	are, respectively, the amounts of primary (α), secondary (β), and final (γ) energy produced using technology j in region k and year t;
$x_\alpha^{kk'}(t), x_\beta^{kk'}(t)$	are, respectively, the (unknown) levels of transportation of primary (α) and secondary (β) energy from region k to region k' in year t;
$a_{\alpha j}^k(t), a_{\beta j}^k(t), a_{\gamma j}^k(t)$	are energy conversion coefficients;
$b_{\alpha j}^k(t), b_{\beta j}^k(t)$	are energy conversion coefficients related to intermediate energy consumption;
$b_\alpha^{kk'}(t), b_\beta^{kk'}(t)$	specify energy losses during transportation; and
$d_\alpha^k(t), d_\beta^k(t), d_\gamma^k(t)$	are, respectively, demands for primary (α), secondary (β), and final (γ) energy in region k and year t.

The constraints on nonenergy resources [referred to later in this report as WELMM factors (Grenon and Lapillone 1976); see also the footnote on p. 27], which are similar to inequality (1.9), are written in the form

$$\sum_{\alpha,k,j} f_{i\alpha j}^k(t) x_{\alpha j}^k(t) + \sum_{\beta,k,j} f_{i\beta j}^k(t) x_{\beta j}^k(t) + \sum_{\gamma,k,j} f_{i\gamma j}^k(t) x_{\gamma j}^k(t) \leqslant f_i(t)$$

For each nonrenewable kind of primary energy α we have a constraint

$$\sum_{j,k,t} x_{\alpha j}^k(t) \leqslant \bar{z}_\alpha$$

which is similar to inequalities (1.31)–(1.33).

It can be seen that these conditions, though much more detailed in form, have the same structure as the constraints of the models discussed earlier. The description of the dynamics of system development differs however in some respects. In the SPI model (A.A. Makarov 1977), the equations linking blocks t and $t + 1$ have the following form

$$\sum_{j \in J_0} x_{ij}(t) + \sum_{j \in J_1} x_{ij}(t) = \tilde{y}_i(t+1) = \sum_{j \in J_0} x_{ij}(t+1) + x_i(t+1)$$

where

i denotes a particular energy unit (plant, power station, etc.); and
j denotes the type of conversion process.

The set of indices J_0 is associated with conversion (or production) capacity which exists at the beginning of period t ("old capacity") and the set of indices J_1 is associated with capacity which was built during period t ("new capacity"); thus $y_i(t+1)$ is the production capacity of type i at the end of year t (or at the beginning of year $t+1$); $x_i(t+1)$ is the capacity of type i which is dismantled in year $t+1$.

The above equations can be rewritten in a form closer to that of the state equation (1.1)

$$\sum_{j \in J_0} x_{ij}(t+1) = \sum_{j \in J_0} x_{ij}(t) + \sum_{j \in J_1} x_{ij}(t) - x_i(t+1)$$

By comparison it is evident that the term $\Sigma_{j \in J_0} x_{ij}(t)$ may be associated with the term $y_i(t)$ in eqn. (1.1), whereas the term $\Sigma_{j \in J_1} x_{ij}(t) - x_i(t+1)$ corresponds to the term $v_i(t) - v_i(t - \tau_i)$ in eqn. (1.1).

The other peculiarity of the SPI model is the objective function. The minimization of the total discounted cost was not considered to be altogether adequate in view of the uncertainty in prices. Therefore, the objective function of the model is given in the form of discounted consumption of total expenditures of different material resources and manpower (WELMM factors)

$$J = \sum_{t=0}^{T-1} \sum_{i} \beta(t) E_i(t) f_i(t)$$

where the coefficient $E_i(t)$ converts the amounts of each resource i into a unified system of units and $\beta(t)$ is a discounting factor.

The dimensions of the SPI model are 500–600 constraints and 4000–5000 variables for the long-range planning variant and 1200–1300 constraints and 6000–7000 variables for the five-year planning problem. To solve these optimization problems a special program package has been developed which gives a three- to four-fold reduction of computation time compared to the conventional simplex method (A.A. Makarov 1977).

2 RESOURCES MODEL

The resources model is designed for the evaluation of long-term resource exploration and extraction strategies. It also provides inputs for the energy supply model (see Section 1), essentially by establishing relations between available quantities of given natural resources and their possible costs of production or extraction (Naill 1972; Brobst and Pratt 1973; Govett and Govett 1974; Kaya and Suzuki 1974; McKelvey 1974; Mesarovic and Pestel 1974; Grenon 1976; Grenon and Lapillone 1976; Grenon and Zimin 1977; Ayres 1978; Kydes 1978).

We will consider the production of natural resources over a given planning horizon at a regional (or national) level. The lengths of each time step and of the whole planning horizon correspond to those in the energy supply model. The availabilities of various resources are expressed in physical units and costs are measured in monetary units.

The model's structure is similar to that of the energy supply model in the sense that it is a DLP model in which the optimal mix of technologies for exploration and extraction of natural energy resources is determined.

2.1 Basic Model

2.1.1 State Equations

The model consists of two subsystems: the resource-accounting subsystem and the capital-stocks subsystem. Using the definitions provided by McKelvey and others (Brobst and Pratt 1973; Govett and Govett 1974; Kaya and Suzuki 1974; McKelvey 1974), the first subsystem describes the movement of resources from the "speculative" to the "hypothetical" category and from the "hypothetical" to the "identified" category. Both renewable and nonrenewable resources may be considered. The second subsystem describes the accumulation and depletion of capacity (capital stocks) for the exploration and extraction of both renewable and nonrenewable resources.

Before continuing with the description of the resource model, let us consider a simple example, which illustrates how the dynamics of the process will be described. Let $x(t)$ be the total amount of nonrenewable resource in place at the beginning of period t. By applying given extraction technologies it is only possible to extract a certain proportion of the total amount of this resource in place. We will denote the extractable (or recoverable) amount of the resource by $\hat{x}(t)$: it is convenient to refer to $\hat{x}(t)$ as a net value and to $x(t)$ as a gross value. The relationship between the gross and net values of the resource may be described by

$$x(t) = \hat{x}(t)/\delta$$

where $\delta (0 < \delta < 1)$ is the recoverability factor of the resource (for a fixed technology) during period t.

Bearing this in mind, we can describe the process in three ways: in terms of gross values, net values, or a mixture of both. Let $u(t)$ be the (gross) amount of the resource extracted in period t, and $\tilde{u}(t)$ be the (gross) amount of the resource moved during the same period from the hypothetical to the identified category. Then the balance equation is

$$x(t+1) = x(t) - u(t) + \tilde{u}(t) \qquad (t = 0, 1, \ldots, T-1)$$

It is evident that

$$x(t) \geqslant 0 \qquad \text{(for all } t)$$

which is equivalent to

$$\sum_{g=0}^{t} u(g) \leqslant x(0) + \sum_{g=0}^{t} \tilde{u}(g) \qquad (t = 1, 2, \ldots, T)$$

To obtain a description in "net" units, all the variables must be multiplied by δ. Due to the linearity of the relationships

$$\hat{x}(t+1) = \hat{x}(t) - \hat{u}(t) + \hat{\tilde{u}}(t)$$

In practice, a mixed description is generally used

$$x(t+1) = x(t) - \hat{u}(t)/\delta(t) + \tilde{u}(t)$$

In this case, the condition

$$x(t) \geqslant 0$$

is equivalent to

$$\sum_{g=0}^{t} u(g) \leqslant \delta\left[x(0) + \sum_{g=0}^{t} \tilde{u}(g)\right]$$

The value

$$x(0) + \sum_{g=0}^{t} \left[\tilde{u}(g) - \hat{u}(g)\right] \geqslant (1-\delta)\left[x(0) + \sum_{g=0}^{t} \tilde{u}(g)\right]$$

denotes the (gross) amount of the resource remaining in place after t periods of extraction.

From this point onwards we will use the mixed description but, for simplicity, we will omit the "hat" sign on variable $\hat{u}(t)$ (Grenon and Zimin 1977).

Nonrenewable resources. Let

- $x_i^1(t)$ be the (gross) amount (or stock) of an identified nonrenewable resource i at period t;
- $u_{mi}^1(t)$ be the (net) amount of resource i extracted by technology m during period t (extraction intensity);
- M_i^1 be the total number of extraction technologies which can be applied to nonrenewable resource i;
- $u_{ki}^2(t)$ be the (gross) amount of resource i moved from the hypothetical to the identified category by exploration technology k during period t; and
- K_i^1 be the total number of exploration technologies which can be applied to nonrenewable resource i.

Then the dynamics (in total amounts) of identified nonrenewable resources will be as follows

$$x_i^1(t+1) = x_i^1(t) - \sum_{m \in M_i^1} u_{mi}^1(t)/\delta_{mi}^1(t) + \sum_{k \in K_i^1} u_{ki}^2(t) \tag{2.1}$$

Here $\delta_{mi}^1(t)$ is the recoverability of resource i by technology m during period t.

For hypothetical resources (all variables are "gross" values) we introduce, in a similar way

- $x_i^2(t)$ as the total amount of resource i in the hypothetical category in period t; and
- $u_i^3(t)$ as the total amount of resource i moved from the speculative to the hypothetical category as a result of exploration activity during period t.

Note that in this case we do not single out different exploration technologies, in contrast to the case of moving resources from the hypothetical to the identified category.

The state equations for this group of hypothetical nonrenewable resources will be as follows

$$x_i^2(t+1) = x_i^2(t) - \sum_{k\in K_i^1} u_{ki}^2(t) + u_i^3(t) \tag{2.2}$$

Similarly, for the speculative category of nonrenewable resources

$$x_i^3(t+1) = x_i^3(t) - u_i^3(t) + u_i^4(t) \tag{2.3}$$

where

- $x_i^3(t)$ is the total estimate of resource i in the speculative category during period t; and
- $u_i^4(t)$ is the change in the estimate of resource i in the speculative category during period t as a result of improved scientific knowledge.

In the state equations (2.1)–(2.3), $\{x_i^1(t), x_i^2(t), x_i^3(t)\}$ $(i = 1, 2, \ldots, N_1)$ are state variables for the nonrenewable resources subsystem, $\{u_{mi}^1(t), u_{ki}^2(t), u_i^3(t), u_i^4(t)\}$ $(m\in M_i^1$, $k\in K_i^1$, $i = 1, \ldots, N_1)$ are control variables, and $i = 1, \ldots, N_1$, where N_1 is the total number of categories of nonrenewable resources considered.

Renewable resources. In a similar way we can write the state equations for renewable resources such as solar, geothermal, etc., as follows

$$y_i^1(t+1) = y_i^1(t) + \sum_{k\in K_i^2} v_{ki}^2(t) \tag{2.4}$$

$$y_i^2(t+1) = y_i^2(t) - \sum_{k\in K_i^2} v_{ki}^2(t) + v_i^3(t) \tag{2.5}$$

$$y_i^3(t+1) = y_i^3(t) - v_i^3(t) + v_i^4(t) \qquad (i = 1, 2, \ldots, N_2) \tag{2.6}$$

where

- $y_i^1(t)$ is the total available flow of renewable resource i in period t;
- $y_i^2(t)$ is the total hypothetical flow of resource i in period t;
- $y_i^3(t)$ is the total speculative flow of resource i in period t;
- $v_{ki}^2(t)$ is the intensity of exploration technology k applied to resource i in period t;
- $v_i^3(t)$ is the total flow of renewable resource i moved from the speculative to the hypothetical category as a result of exploration activity during period t;
- $v_i^4(t)$ is the change in the estimated flow of renewable resource i in the speculative category during period t as a result of improved scientific knowledge;
- K_i^2 is the total number of exploration technologies for resource i; and
- N_2 is the total number of categories of renewable resources considered.

In the renewable-resources subsystem (2.4)–(2.6), $\{y_i^1(t), y_i^2(t), y_i^3(t)\}$ $(i = 1, 2, \ldots, N_2)$ are the state variables, and $\{v_{ki}^2(t), v_i^3(t), v_i^4(t)\}$ $(k = 1, 2, \ldots, K_i^2;\ i = 1, 2, \ldots, N_2)$ are the control variables.

Initial conditions are assumed to be given for all resource categories

$$\left.\begin{array}{ll} x_i^1(0) = x_i^{1,0};\ x_i^2(0) = x_i^{2,0};\ x_i^3(0) = x_i^{3,0} & (i = 1, 2, \ldots, N_1) \\ y_i^1(0) = y_i^{1,0};\ y_i^2(0) = y_i^{2,0};\ y_i^3(0) = y_i^{3,0} & (i = 1, 2, \ldots, N_2) \end{array}\right\} \tag{2.7}$$

Dynamics of extraction and exploration capacity. Alongside the subsystems which describe resource extraction and exploration themselves, it is necessary to introduce a subsystem describing the *development* of resource extraction and exploration capacity. This can be done by using equations similar to eqn. (1.1). For the extraction part of the subsystem, let

$z_m(t)$ be the extraction capacity of type m in period t;
$w_m(t)$ the increase of the mth extraction capacity during period t; and
τ_m the service lifetime of units of capacity of type m.

Then the state equations for this submodel will be as follows

$$z_m(t+1) = z_m(t) + w_m(t) - w_m(t - \tau_m) \tag{2.8}$$

where, in the general case, $m \in M_1 \cup M_2$, the union of two sets

M_1 (the total set of technologies for extracting nonrenewable resources); and
M_2 (the total set of technologies for extracting renewable resources.)

Initial conditions are given as follows

$$z_m(0) = z_m^0 \tag{2.9a}$$

$$w_m(t - \tau_m) = w_m^0(t - \tau_m) \qquad (0 \leqslant t \leqslant \tau_m - 1) \tag{2.9b}$$

The dynamics of the development of exploration capacity can be described in a similar way, but for simplicity these equations are omitted here.

2.1.2 Constraints

The activities of exploration and extraction of natural resources are subject to a number of constraints. In the sections which follow we will examine how the model deals with physical, recoverability, availability, and demand constraints.

Physical sense. By virtue of their physical meaning, all the variables in the model are nonnegative

$$\left.\begin{array}{l} x_i^1(t) \geqslant 0;\ x_i^2(t) \geqslant 0;\ x_i^3(t) \geqslant 0 \\ u_{mi}^1(t) \geqslant 0;\ u_{ki}^2(t) \geqslant 0;\ u_i^3(t) \geqslant 0;\ u_i^4(t) \geqslant 0 \\ \qquad (i = 1, 2, \ldots, N_1;\ m = 1, 2, \ldots, M_1;\ k = 1, 2, \ldots, K_1) \end{array}\right\} \tag{2.10}$$

$$\left.\begin{array}{l} y_i^1(t) \geqslant 0;\ y_i^2(t) \geqslant 0;\ y_i^3(t) \geqslant 0 \\ v_{mi}^1(t) \geqslant 0;\ v_{ki}^2(t) \geqslant 0;\ v_i^3(t) \geqslant 0;\ v_i^4(t) \geqslant 0 \\ z_m(t) \geqslant 0;\ w_m(t) \geqslant 0;\ m \in M_1 \cup M_2 \\ \qquad (i = 1, 2, \ldots, N_2;\ m = 1, 2, \ldots, M_2;\ k = 1, 2, \ldots, K_2) \end{array}\right\} \tag{2.11}$$

Recoverability. The recoverability of a resource is assumed to be associated with the type of resource and the technology used for its extraction. As mentioned previously, the nonnegativity condition for nonrenewable resources may be stated as

$$x_i^1(t) \geqslant 0 \tag{2.12}$$

which [from eqns. (2.1) and (2.7)] is equivalent to

$$\sum_{g=0}^{t} \sum_{m \in M_i^1} u_{mi}^1(g)/\delta_{mi}^1(g) \leqslant x_i^{1,0} + \sum_{g=0}^{t} \sum_{k \in K_i^1} u_{ki}^2(g) \qquad (i = 1, 2, \ldots, N_1) \tag{2.12a}$$

For renewable resources the corresponding constraints may be written as

$$\sum_{m \in M_i^2} v_{mi}^1(t)/\delta_{mi}^2(t) \leqslant y_i^1(t) \qquad (i = 1, 2, \ldots, N_2) \tag{2.13}$$

Here $v_{mi}^1(t)$ is the amount of the renewable resource i utilized by technology $m \in M_i^2$ during period t (the "extraction" intensity). In contrast to eqn. (2.1), this variable does not enter eqn. (2.4) for renewable resources, because utilization of such resources (solar, geothermal, etc.) does not influence their source.

From eqns. (2.4) and (2.7), condition (2.13) is equivalent to

$$\sum_{m \in M_i^2} v_{mi}^1(t)/\delta_{mi}^2(t) \leqslant y_i^{1,0} + \sum_{g=0}^{t-1} \sum_{k \in K_i^2} v_{ki}^2(g) \tag{2.14}$$

Availability. In their simplest form, these constraints can be expressed as upper bounds on control variables

$$u_{mi}^1(t) \leqslant \bar{u}_{mi}^1(t);\ u_{ki}^2(t) \leqslant \bar{u}_{ki}^2(t);\ u_i^3(t) \leqslant \bar{u}_i^3(t);\ u_i^4(t) \leqslant \bar{u}_i^4(t) \tag{2.15}$$

and

$$v_{mi}^1(t) \leqslant \bar{v}_{mi}^1(t);\ v_{ki}^2(t) \leqslant \bar{v}_{ki}^2(t);\ v_i^3(t) \leqslant \bar{v}_i^3(t);\ v_i^4(t) \leqslant \bar{v}_i^4(t) \tag{2.16}$$

These constraints are similar to those of inequality (1.8), and express very approximately the availability over time of various technologies for exploration and extraction.

The development of a given resource system may often require the input of other resources (such as land, manpower, etc.) which are external to the system itself (referred to here as WELMM* factors). These constraints can be written in a form similar to that of inequality (1.9)

$$\sum_{s,i} r_{si}^{vlu}(t) u_{si}^{v}(t) \leq R^{vlu}(t) \tag{2.17}$$

$$\sum_{q,i} r_{qi}^{vlv}(t) v_{qi}^{v}(t) \leq R^{vlv}(t) \qquad (l = 1, 2, \ldots, L; v = 1,2,3,4) \tag{2.18}$$

where

$R^{vlu}(t), R^{vlv}(t)$ are, respectively, the amounts of nonrenewable and renewable external resource l (or WELMM factor l), available in period t for each group of exploration activities v;

L is the total number of WELMM factors considered as external to the model; and

$r_{si}^{vlu}(t), r_{qi}^{vlv}$ are, respectively, the (normative) consumptions of nonrenewable and renewable WELMM factor l per unit of productive output; and

$$s \in M_i^1, \text{ if } v = 1; \; s \in K_i^1, \text{ if } v = 2$$

$$q \in M_i^2, \text{ if } v = 1; \; q \in K_i^2, \text{ if } v = 2$$

The subscripts s and q on the left-hand sides of inequalities (2.17) and (2.18) should be dropped if $v = 3$ or 4. In practical terms, coefficients $r_{si}^{vlu}(t)$ and $r_{qi}^{vlv}(t)$ are negligibly small for $v = 2, 3$, or 4.

The other important type of availability constraint is connected with the linkage of resource-extraction and production capacity: the extraction of resources during each period is limited by the production capacity available

$$\sum_i u_{mi}^1(t) \leq z_m(t) \qquad (m \in M_1) \tag{2.19}$$

$$\sum_i v_{mi}^1(t) \leq z_m(t) \qquad (m \in M_2) \tag{2.20}$$

where $z_m(t)$, $m \in M_1$, and $m \in M_2$ are defined from eqn. (2.8).

In its turn, the development of the extraction-capacity subsystem (2.8) may itself be limited by the amount of resources available for construction of new capacity. In this case, the control variables $w_m(t)$ in eqn. (2.8) are subject to constraints which are similar to those described in inequalities (2.17) and (2.18).

---

* Grenon and Lapillone (1976) originally used WELMM as an abbreviation for Water, Energy, Land, Materials, and Manpower; however in this report we use the term "WELMM factor" to mean any arbitrary resource which is external to the system in question.

Demand. Demands are exogeneous for the resource model. These constraints can be written in the form

$$\sum_{m\in M_i^1} u_{mi}^1(t) \geqslant d_i^u(t) \qquad (i=1,2,\ldots,N_1) \tag{2.21}$$

for nonrenewable resources, and in the form

$$\sum_{m\in M_i^2} v_{mi}^1(t) \geqslant d_i^v(t) \qquad (i=1,2,\ldots,N_2) \tag{2.22}$$

for renewable resources, where $d_i^u(t)$ and $d_i^v(t)$ are, respectively, the demands for nonrenewable and renewable resource i in period t.

It should be noted that accurate estimation of the demands $d_i^u(t)$ and $d_i^v(t)$ is very important in the resource model: this is because these parameters exert a strong influence on the timing and corresponding costs of putting into operation new extraction technologies and on the intensity of exploration activities, and therefore, finally, on the optimal solution itself.

2.1.3 Objective Function

A variety of different objective functions is possible for the resource system development. Following the ESS model procedure, we define the objective function so as to minimize the total discounted costs required to implement a given resource-development strategy

$$\begin{aligned}
&J(u^1,u^2,u^3,u^4,v^1,v^2,v^3,v^4,w) = \\
&\sum_{t=0}^{T-1}\beta(t)\Bigg\{\Bigg[\sum_{m,i} c_{mi}^{1u}u_{mi}^1(t) + \sum_{k,i} c_{ki}^{2u}u_{ki}^2(t) + \sum_i c_i^{3u}u_i^3(t) + \sum_i c_i^{4u}u_i^4(t)\Bigg] \\
&+\Bigg[\sum_{m,i} c_{mi}^{1v}v_{mi}^1(t) + \sum_{k,i} c_{ki}^{2v}v_{ki}^2(t) + \sum_i c_i^{3v}v_i^3(t) + \sum_i c_i^{4v}v_i^4(t)\Bigg] + \sum_m c_m^z z_m(t) + \sum_m c_m^w w_m(t) \\
&+\Bigg[\sum_{l,m,i} c_{mi}^{1lu}r_{mi}^{1lu}u_{mi}^1(t) + \sum_{l,k,i} c_{ki}^{2lu}r_{ki}^{2lu}u_{ki}^2(t) + \sum_{l,i} c_i^{3lu}r_i^{3lu}u_i^3(t) + \sum_{l,i} c_i^{4lu}r_i^{4lu}u_i^4(t)\Bigg] \\
&+\Bigg[\sum_{l,m,i} c_{mi}^{1lv}r_{mi}^{1lv}v_{mi}^1(t) + \sum_{l,k,i} c_{ki}^{2lv}r_{ki}^{2lv}v_{ki}^2(t) + \sum_{l,i} c_i^{3lv}r_i^{3lv}v_i^3(t) + \sum_{l,i} c_i^{4lv}r_i^{4lv}v_i^4(t)\Bigg]\Bigg\}
\end{aligned}$$

Here

$c_{mi}^{1u}, c_{ki}^{2u}, c_i^{3u}, c_i^{4u}$	are exploration costs for nonrenewable resources;
$c_{mi}^{1v}, c_{ki}^{2v}, c_i^{3v}, c_i^{4v}$	are exploration costs for renewable resources;
c_m^z	are operational costs;
c_m^w	are capital investment costs; and
$c_{mi}^{1lu}, c_{mi}^{1lv}$, etc.	are costs of WELMM factors (external resources).

Transportation costs can also be included in the model.

2.1.4 Statement of the Problem

Finally we can formulate the problem of optimal development of the resource system as follows.

Problem 2.1. Given the state equations for the nonrenewable resources subsystem $(i = 1, 2, \ldots, N_1)$

$$x_i^1(t+1) = x_i^1(t) - \sum_{m \in M_i^1} u_{mi}^1(t)/\delta_{mi}^1(t) + \sum_{k \in K_i^1} u_{ki}^2(t);\ x_i^1(0) = x_i^{1,0}$$

$$x_i^2(t+1) = x_i^2(t) - \sum_{k \in K_i^1} u_{ki}^2(t) + u_i^3(t);\ x_i^2(0) = x_i^{2,0}$$

$$x_i^3(t+1) = x_i^3(t) - u_i^3(t) + u_i^4(t);\ x_i^3(0) = x_i^{3,0}$$

for the renewable resources subsystem $(i = 1, 2, \ldots, N_2)$

$$y_i^1(t+1) = y_i^1(t) + \sum_{k \in K_i^2} v_{ki}^2(t);\ y_i^1(0) = y_i^{1,0}$$

$$y_i^2(t+1) = y_i^2(t) - \sum_{k \in K_i^2} v_{ki}^2(t) + v_i^3(t);\ y_i^2(0) = y_i^{2,0}$$

$$y_i^3(t+1) = y_i^3(t) - v_i^3(t) + v_i^4(t);\ y_i^3(0) = y_i^{3,0}$$

and for the extraction capacity subsystem ($m \in \{1, \ldots, M_1\}$ and $m \in \{1, \ldots, M_2\}$)

$$z_m(t+1) = z_m(t) + w_m(t) - w_m(t - \tau_m);\ z_m(0) = z_m^0$$

$$w_m(t - \tau_m) = w_m^0(t - \tau_m) \qquad (0 \leqslant t \leqslant \tau_m - 1)$$

find controls $\{u_{mi}^1(t), u_{ki}^2(t), u_i^3(t), u_i^4(t)\}$, $\{v_{mi}^1(t), v_{ki}^2(t), v_i^3(t), v_i^4(t)\}$, and $\{w_m(t)\}$, and corresponding trajectories $\{x_i^1(t), x_i^2(t), x_i^3(t)\}$, $\{y_i^1(t), y_i^2(t), y_i^3(t)\}$, and $\{z_m(t)\}$, which satisfy the following constraints

(a) nonnegativity

$$u_{mi}^1(t) \geqslant 0;\ u_{ki}^2(t) \geqslant 0;\ u_i^3(t) \geqslant 0;\ u_i^4(t) \geqslant 0$$

$$v_{mi}^1(t) \geqslant 0;\ v_{ki}^2(t) \geqslant 0;\ v_i^3(t) \geqslant 0;\ v_i^4(t) \geqslant 0$$

$$x_i^1(t) \geqslant 0;\ x_i^2(t) \geqslant 0;\ x_i^3(t) \geqslant 0$$

$$y_i^1(t) \geqslant 0;\ y_i^2(t) \geqslant 0;\ y_i^3(t) \geqslant 0$$

(b) recoverability

$$x_i^1(t) \geqslant 0 \qquad (i = 1, 2, \ldots, N_1)$$

$$\sum_m v_{mi}^1(t)/\delta_{mi}^2(t) \leqslant y_i^1(t) \qquad (i = 1, 2, \ldots, N_2)$$

(c) external-resource availability

$$\sum_{s,i} r_{si}^{vlu}(t) u_{si}^{v}(t) \leqslant R^{vlu}(t)$$

$$(l = 1, 2, \ldots, L;\ v = 1, 2, 3, 4)$$

$$\sum_{q,i} r_{qi}^{vlv}(t) v_{qi}^{v}(t) \leqslant R^{vlv}(t)$$

(d) production-capacity availability

$$\sum_{i} u_{mi}^{1}(t) \leqslant z_m(t) \qquad (m \in M_1)$$

$$\sum_{i} v_{mi}^{1}(t) \leqslant z_m(t) \qquad (m \in M_2)$$

(e) demand

$$\sum_{m} u_{mi}^{1}(t) \geqslant d_i^{u}(t) \qquad (i = 1, 2, \ldots, N_1)$$

$$\sum_{m} v_{mi}^{1}(t) \geqslant d_i^{v}(t) \qquad (i = 1, 2, \ldots, N_2)$$

and minimize the objective function ($v = 1, 2, 3, 4$)

$$J(u, v, w) = \sum_{t=0}^{T-1} \beta(t) \Bigg[\sum_{v,s,i} c_{si}^{vu} u_{si}^{v}(t) + \sum_{v,s,i} c_{si}^{vv} v_{si}^{v}(t) + \sum_{m} c_m^{z} z_m(t) + \sum_{m} c_m^{w} w_m(t) + \sum_{v,l,s,i} c_{si}^{vlu} r_{si}^{vlu} u_{si}^{v}(t) + \sum_{v,l,s,i} c_{si}^{vlv} r_{si}^{vlv} v_{si}^{v}(t) \Bigg]$$

This particular objective function is given here only for illustration. Many other objectives, for instance, the minimization of the total production costs of primary energy resources and effect of their use in the energy sector, are of practical interest, and some examples of such modifications of the model are given in the next section.

2.2 Discussion

The formulation of Problem 2.1 is general enough to allow different modifications to the basic problem. These modifications make it possible to carry out policy analyses for extraction and/or exploration activities, for a single resource or for a group of resources, for a region or a country; it is also possible to determine optimal balances of these activities for nonrenewable and renewable resources. We will now consider some examples of these modifications and particular cases of Problem 2.1.

2.2.1 Extraction and Exploration Model

First we consider the analysis of the interrelationships between extraction and exploration activities for a given nonrenewable energy resource (e.g., coal, oil, etc.).

The problem is as follows. For a given region (or country) there are known initial values for identified and hypothetical stocks of the resource, classified in n different categories (e.g., onshore crude oil, natural gas, and offshore crude oil). There are also M different extraction and K different exploration technologies. The degree of utilization of these technologies depends, during a given period, on the extraction and exploration capacity available during the same period. The problem is to determine the optimal mix of extraction and exploration activities over a given planning horizon which is, at the same time, balanced with the development of the exploration-, extraction-, and production-capacity subsystems and yields the maximum output over the same horizon.

Using the conditions of Problem 2.1, this problem can be formalized as follows.

Problem 2.2 Let the initial stocks of identified and hypothetical resources be given, respectively, as

$$x_i^1(0) = x_i^{1,0} \text{ and } x_i^2(0) = x_i^{2,0} \tag{2.23}$$

with state equations for extraction activities

$$x_i^1(t+1) = x_i^1(t) - \sum_{m \in M_i} u_{mi}^1(t)/\delta_{mi}^1(t) + \sum_{k \in K_i} u_{ki}^2(t) \tag{2.24}$$

and for exploration activities

$$x_i^2(t+1) = x_i^2(t) - \sum_{k \in K_i} u_{ki}^2(t) + \tilde{u}_i^2(t) \tag{2.25}$$

where $\tilde{u}_i^2(t)$ is the increase in the hypothetical stocks of resource i during period t (the discovery rate). In addition, let the initial values of the extraction and exploration capacities be given, respectively, by

$$z_m^1(0) = z_m^{1,0} \text{ and } z_k^2(0) = z_k^{2,0} \tag{2.26}$$

with the state equations

$$z_m^1(t+1) = z_m^1(t) + w_m^1(t) - w_m^1(t - \tau_m^1) \tag{2.27}$$

$$z_k^2(t+1) = z_k^2(t) + w_k^2(t) - w_k^2(t - \tau_k^2) \tag{2.28}$$

The intensities of extraction and exploration activities, $u_{mi}^1(t)$ and $u_{ki}^2(t)$, as well as the intensities of construction of new extraction and exploration capacity, $w_i^1(t)$ and $w_i^2(t)$, are subject to budgetary and other resource constraints

$$\sum_{m,i} r_{mil}^{1u}(t) u_{mi}^1(t) + \sum_{k,i} r_{kil}^{2u}(t) u_{ki}^2(t) + \sum_i r_{il}^{1w}(t) w_i^1(t) + \sum_i r_{il}^{2w}(t) w_i^2(t) \leqslant R_l(t) \tag{2.29}$$

$$\sum_i u_{mi}^1(t) \leqslant z_m^1(t); \ \sum_i u_{ki}^2(t) \leqslant z_k^2(t) \tag{2.30}$$

$$x_i^1(t) \geqslant 0 \tag{2.31}$$

Find nonnegative control sequences $\{u_{mi}^1(t)\}$, $\{u_{ki}^2(t)\}$, and $\{w_i^1(t)\}$, $\{w_i^2(t)\}$, and corresponding nonnegative state variables $\{x_i^1(t)\}$, $\{x_i^2(t)\}$, and $\{z_i^1(t)\}$, $\{z_i^2(t)\}$, which maximize the total output of resource i

$$J = \sum_{t=0}^{T-1} \sum_{m,i} \kappa_i u_{mi}^1(t) \tag{2.32}$$

where κ_i is the energy conversion factor for resource i. Here $\tilde{u}_i^2(t)$ (the discovery rate) is considered as a scenario variable.

2.2.2 Extraction Model

If the increase $\{\tilde{u}_i(t)\}$ of the identified resource is considered as a scenario variable (but not as a result of controllable exploration activities), then the state equations for the extraction system are simplified

$$x_i(t+1) = x_i(t) - u_i(t)/\delta_i(t) + \tilde{u}_i(t); \; x_i(0) = x_i^0 \tag{2.33}$$

where $\tilde{u}_i(t)$ is the amount of resource i moved from the hypothetical to the identified category during period t, and $u_i(t)$ is the total amount of resource i extracted during period t (in this example, different extraction technologies are not singled out).

The development of the extraction capacity subsystem is described by a state equation similar to eqn. (2.27)

$$z_i(t+1) = z_i(t) + w_i(t) - w_i(t - \tau_i); \; z_i(0) = z_i^0 \tag{2.34}$$

with the constraints

$$u_i(t) \leqslant z_i(t); \; u_i(t) \geqslant 0; \; w_i(t) \geqslant 0; \; z_i(t) \geqslant 0 \tag{2.35}$$

$$\sum_i r_{il}^w(t) w_i(t) + \sum_i r_{il}^u(t) u_i(t) \leqslant R_l(t); \; w_i(t) \geqslant 0 \tag{2.36}$$

$$x_i(t) \geqslant 0 \tag{2.36a}$$

The problem is to determine the extraction policy for a given identified resource, subject to constraints on extraction capacity (2.35), availability of external resources (2.36), and recoverability of the given resource (2.36a), which gives the maximum total output during the planning period.

The objective function may be written again as (2.32), or, if we introduce $\xi(t)$ as the cumulative amount of the resource extracted

$$\xi(t+1) = \xi(t) + \sum_i \kappa_i u_i(t); \; \xi(0) = 0 \tag{2.37}$$

as the maximization of $\xi(T)$.

2.2.3 Exploration Model

This model allows us to determine those exploration policies which will move the maximum amount of resources from the hypothetical to the identified category. The subsystem is a counterpart of the extraction subsystem and is described by the equations

$$x_i(t+1) = x_i(t) - u_i(t) + \tilde{u}_i(t);\; x_i(0) = x_i^0 \tag{2.38}$$

$$z_i(t+1) = z_i(t) + w_i(t) - w_i(t-\tau_i);\; z_i(0) = z_i^0 \tag{2.39}$$

$$u_i(t) \leqslant z_i(t);\; u_i(t) \geqslant 0,\; z_i(t) \geqslant 0 \tag{2.40}$$

$$\sum_i r_{il}(t) w_i(t) \leqslant R_l(t);\; w_i(t) \geqslant 0 \tag{2.41}$$

$$x_i(t) \geqslant 0 \tag{2.42}$$

$$J = \sum_{t=0}^{T-1} \sum_i u_i(t) \to \max \tag{2.43}$$

2.2.4 Cost Minimization

In the examples above the objective was to maximize the output from the extraction and/or the exploration subsystems. For many practical purposes it is also necessary to calculate the relationship between the optimal cost J^* and the cumulative availability of a given resource [for example, for calculating cost coefficients in the objective function (1.12) of the energy supply system model]. This can be done by using a simple optimization model

$$\left.\begin{aligned}
&x_i(t+1) = x_i(t) - u_i(t)/\delta_i(t) + \tilde{u}_i(t);\; x_i(0) = x_i^0\\
&z_i(t+1) = z_i(t) + w_i(t) - w_i(t-\tau_i);\; z_i(0) = z_i^0\\
&\xi(t+1) = \xi(t) + \sum_i \kappa_i(t) u_i(t);\; \xi(0) = 0\\
&\sum_i \kappa_i u_i(t) \geqslant d(t);\; u_i(t) \geqslant 0\\
&u_i(t) \leqslant z_i(t);\; z_i(t) \geqslant 0\\
&x_i(t) \geqslant 0
\end{aligned}\right\} \tag{2.44}$$

$$J = \sum_t \sum_i [c_i^u(t) u_i(t) + c_i^w(t) w_i(t)] \to \min \tag{2.45}$$

This model differs from the extraction model in two ways: demand constraints are included (2.44), and the objective function (2.45) is formulated differently. Resource constraints (2.36) are omitted here because they are implicitly accounted for by cost coefficients $c_i^u(t)$ and $c_i^w(t)$ in objective function (2.45).

Clearly, in this simple model

$$\sum_i \kappa_i u_i^*(t) = d(t)$$

for optimal $u_i^*(t)$. Hence

$$\xi(t+1) = \xi(t) + d(t);\; \xi(0) = 0 \tag{2.46}$$

and

$$\xi(T) = \sum_{t=0}^{T-1} d(t) \tag{2.47}$$

The problem is, therefore, to calculate cost–supply curves

$$J^* = J[u^*, z(T)] = \varphi[z(T)]$$

It should be noted that the behavior of these curves is strongly dependent on the behavior of the demand curve $d(t)$.

2.2.5 Dimensions of the Models

Finally, we will calculate the typical dimensions of the resources model. Let

- M be the total number of different countries in a region;
- L be the number of resource provinces within a country;
- K be the number of basins within a province;
- T be the length of the planning horizon;
- l be the number of different resource categories in a basin;
- m be the number of different technologies which can be used in exploration and extraction; and
- k be the number of WELMM factors limiting extraction.

One can see that the model will have a total of $(3l + m)KLM$ state equations, $(2l + k + m)$ KLM constraints (nonnegativity constraints are not included here), and $3lmKLM$ control variables for each period.

For example, consider a region consisting of only one country with two resource provinces. Assuming that the average number of basins in a province is three, the average number of different resource categories is two (for instance, crude oil and natural gas), the number of different technologies is two, and the number of limiting WELMM factors is two, we calculate that, for each period, the model would have 48 state equations, 48 constraints, and 72 control variables. Thus, for a problem of quite realistic size, the resources model is manageable and can be handled even by standard LP-solving programs.

2.2.6 Resource Modeling under Conditions of Uncertainty

One of the intrinsic features of the resources model is uncertainty in the values of various parameters, particularly for the speculative and hypothetical resource categories. The conventional method for handling this difficulty is to consider these parameters as scenario variables [e.g., $\tilde{u}_i^2(t)$ in eqn. (2.25), or $\tilde{u}_i(t)$ in eqn. (2.33)], carrying out numerous computer runs for different hypothetical values of the variables.

A more sophisticated approach is to consider "maxmin" problems associated with the given model. The maxmin approach allows us to evaluate upper and lower limits of the objective function for optimization problems under conditions of uncertainty, and to elaborate extraction and exploration policies which guarantee the required results within a given range of uncertain parameters. Methods for solving maxmin DLP problems have been considered by Propoi and Yadykin (1974).

Yet another approach to the treatment of uncertainty conditions in resource models is the statement of the problem in a multistage stochastic programming framework (Ermoliev 1978).

3 ECONOMIC DEVELOPMENT MODELS

In this section we present a model which simulates optimal behavior of the entire economy of a given region for various different objectives. Interest in such models has been increasing in recent years because they allow us to calculate various "optimal" mixes of the dynamics of such important economic indicators as production levels, capital investment, and levels of intermediate and final consumption of goods produced. A number of different optimization models of economic development have been described previously (see, for example, Kantorovich 1965; V.V. Makarov 1966; Ivanilov and Petrov 1970a, b; Aganbegyan et al. 1974; Aganbegyan and Valtukh 1975). However, we will not analyze all these models here, but will restrict ourselves to describing a multibranch industrial model named INTERLINK (Zimin 1976a, b, 1977, 1980), which is conceptually based on its predecessor, the π-model developed at the Computer Center of the USSR Academy of Sciences (Ivanilov and Petrov 1970a, b). The model presented below may be viewed as a simplified version of the original π-model.

3.1 Basic Model

3.1.1 State Equations

The system under consideration is broken down into two subsystems, describing production and the development of capacity (or capital stock accumulation).

Production subsystem. The operation of industry is described in terms of n producing sectors. Let

- $x_i(t)$ be the cumulative production in sector i ($i = 1,2, \ldots, n$) up to period t;
- $u_i(t)$ be the gross output (production level) of sector i during period t;
- $v_i(t)$ be the additional capital stock (plant, equipment, etc.) constructed in period t; and
- $a_{ij}(t)$ be the input–output coefficients (i.e. the number of units of i required to produce one unit of j).

In addition, we assume that

- τ_j is the time (number of periods) required to construct and put into operation additional capacity in sector j;
- $b_{ij}(\tau)$ are capital coefficients representing the amount of sector i products required to build unit capacity in sector j, to be available for production τ periods later;
- $w_i(t)$ is the final consumption of sector i products during period t; and
- $s_i(t)$ is the net amount of sector i products exported during period t.

Then the state equations describing the production subsystem can be written as follows

$$x_i(t+1) = x_i(t) + u_i(t) - \sum_{j=1}^{n} a_{ij}(t)u_j(t) - \sum_{j=1}^{n}\sum_{\tau=0}^{\tau_j} b_{ij}(\tau)v_j(t-\tau) - w_i(t) - s_i(t)$$

$$(i = 1, 2, \ldots, n;\ t = 0, 1, \ldots, T-1) \tag{3.1}$$

Initial inventories and preplanning controls are assumed to be given by

$$x_i(0) = x_i^0 \qquad (i = 1, 2, \ldots, n;\ t = 0, 1, \ldots, \tau_i - 1) \tag{3.2}$$

$$v_i(t-\tau_i) = v_i^0(t-\tau_i) \qquad (i = 1, 2, \ldots, n;\ t = 0, 1, \ldots, \tau_i - 1) \tag{3.3}$$

Assuming that $\tau_j = \bar{\tau}$ for all sectors $j (j = 1, 2, \ldots, n)$ eqn. (3.1) can be rewritten in matrix form

$$x(t+1) = x(t) + (\mathrm{I} - \mathrm{A}(t))u(t) - \sum_{\tau=0}^{\bar{\tau}} \mathrm{B}(\tau)v(t-\tau) - w(t) - s(t) \tag{3.1a}$$

where

$x(t) = \{x_i(t)\}$ is a state vector, $u(t) = \{u_i(t)\}$;
$v(t) = \{v_i(t)\}, w(t) = \{w_i(t)\}$ are control vectors; and
$s(t) = \{s_i(t)\}$ is considered here as an exogenous vector.

For some particular problems, the export/import variables must be considered as control (or decision) variables. In these cases the net export $s(t)$ is better represented as follows

$$s(t) = s^+(t) - s^-(t) \qquad (s^+(t) \geqslant 0, s^-(t) \geqslant 0)$$

where $s^+(t)$ is the import vector and $s^-(t)$ is the export vector.

Development of capacity subsystem. Let

$y_i(t)$ be the production capacity in sector i $(i = 1, 2, \ldots, n)$ at time t; and
$d_i(t)$ be the depreciation factor in sector i during period t.

Then the dynamics of production capacity may be written as follows

$$y_i(t+1) = (1 - d_i(t))y_i(t) + v_i(t-\tau_i) \qquad (i = 1, 2, \ldots, n) \tag{3.4}$$

The initial capital stocks (plant, equipment, etc.) are given as

$$y_i(0) = y_i^0 \tag{3.5}$$

Assuming again for simplicity that

$$\tau_i = \bar{\tau} \qquad \text{(for all } i\text{)}$$

we can rewrite eqn. (3.4) in matrix form

$$y(t+1) = [\mathbf{I} - \mathbf{D}(t)]\,y(t) + v(t - \bar{\tau}) \tag{3.4a}$$

where $\mathbf{D}(t)$ is a diagonal matrix with $d_i(t)$ on the main diagonal, and $y(t) = \{y_i(t)\}$ $(i = 1, 2, \ldots, n)$ is a state vector for the production capacity subsystem.

3.1.2 Constraints

It is evident that any economic system operates within certain constraints; this implies a range of physical, economic, institutional, and other limits to our choice of the control variables which we will use in the model.

Physical Sense. All state and control variables are nonnegative

$$\left.\begin{array}{c} u_i(t) \geqslant 0;\; v_i(t) \geqslant 0;\; w_i(t) \geqslant 0;\; x_i(t) \geqslant 0;\; y_i(t) \geqslant 0 \\ (i = 1, 2, \ldots, n;\; t = 0, 1, \ldots, T-1) \end{array}\right\} \tag{3.6}$$

Resource availability. The production system requires certain external resource inputs for its operation. At their most basic, these are inputs of labor and primary resources. Both constraints can be written in a similar way

(a) for labor resources

$$\sum_{j=1}^{n} l_{kj}(t) u_j(t) \leqslant l_k(t) \qquad (k = 1, 2, \ldots, K) \tag{3.7}$$

where

$l_k(t)$ is the total labor of category $k (k = 1, 2, \ldots, K)$ available in period t; and
$l_{kj}(t)$ are the labor output ratios for sector j.

(b) for other primary resources (described here as WELMM factors)

$$\sum_{j=1}^{n} r_{mj}(t) u_j(t) \leqslant r_m(t) \qquad (m = 1, 2, \ldots, M) \tag{3.8}$$

where

$r_m(t)$ is the total amount of resource category m (WELMM factor m) available during period t; and
$r_{mj}(t)$ are specific resource requirements per unit of sector j production (resource–output ratios) during period t.

In matrix form, inequalities (3.7) and (3.8) become

$$\mathbf{L}(t)u(t) \leqslant l(t) \tag{3.7a}$$

$$\mathbf{R}(t)u(t) \leqslant r(t) \tag{3.8a}$$

Production capacity. The gross output of each sector is limited by the available production capacity in that sector

$$u_i(t) \leqslant y_i(t) \qquad (i = 1, 2, \ldots, n) \tag{3.9}$$

or, in vector form

$$u(t) \leqslant y(t) \tag{3.9a}$$

Inventory. These constraints relate to the possibility of accumulating limited stocks of a given commodity*. For storable goods

$$0 \leqslant x_i(t) \leqslant \bar{x}_i(t) \tag{3.10}$$

where

$\bar{x}_i(t)$ are the given stock capacities; and
$x_i(t)$ are calculated from eqn. (3.1).

For nonstorable goods we write, instead of inequality (3.10)

$$u_i(t) - \sum_{j=1}^{n} a_{ij}(t)u_j(t) - \sum_{j=1}^{n} \sum_{T=0}^{T_j} b_{ij}(T)v_j(t - T) - w_i(t) - s_i(t) \geqslant 0 \tag{3.11}$$

or, in matrix form

$$[\mathbf{I} - \mathbf{A}(t)]\, u(t) - \sum_{T=0}^{\bar{T}} \mathbf{B}(T)v(t - T) - w(t) - s(t) \geqslant 0 \tag{3.12}$$

It should be stressed that, in many practical cases, the accumulation of large stocks of goods is either physically unreasonable or prohibitively expensive. Hence, $\{x_i(t)\}$ values are small in comparison to the outputs of the system. Therefore we can consider the balance equation (or bill of goods) in the form of an inequality [equivalent to inequality (3.12)]

$$[\mathbf{I} - \mathbf{A}(t)]\, u(t) \geqslant \sum_{T=0}^{\bar{T}} \mathbf{B}(T)v(t - T) + w(t) + s(t) \tag{3.13}$$

or as an equation

$$[\mathbf{I} - \mathbf{A}(t)]\, u(t) = \sum_{T=0}^{\bar{T}} \mathbf{B}(T)v(t - T) + w(t) + s(t) \tag{3.13a}$$

for both storable and nonstorable goods.

Consumption. Final consumption usually has limits for each sector *i*. In many cases it can be represented by an inequality of the form

*In addition, note that here we regard such resources as manpower and electricity as nonstorable goods.

$$w_i(t) \geqslant g_i(t)\omega(t) \tag{3.14}$$

where

$\omega(t)$ is the total final consumption of all goods; and
$g_i(t)$ is the share of total consumption provided by sector i.

The exogenously-given vector $g(t) = \{g_i(t)\}$ $(i = 1, 2, \ldots, n)$ predefines the profile of final consumption over time. The introduction of a consumption profile allows one to use a scalar control $\omega(t)$ instead of the control vector $w(t)$

$$w(t) \geqslant g(t)\omega(t) \tag{3.14a}$$

3.1.3 Objective Function

In the sections above, $\{u,v,w\} = \{u_i(t), v_i(t), w_i(t)\}$ are control variables, and $\{x,y\} = \{x_i(t), y_i(t)\}$ are state variables. The choice of optimal controls depends on the choice of the objective function for a particular problem. We will now consider typical examples of the objective function.

Maximization of the cumulative discounted-goods supply. In this case, the objective function (in monetary terms) is

$$J = \sum_{t=0}^{T-1} \beta(t)\omega(t) \tag{3.15}$$

where $\beta(t)$ is the discounting factor. If we consider only the last step of the planning horizon then the objective function (in terms of products) will be

$$J = \sum_{i=1}^{n} h_i^w(T)w_i(T) \tag{3.16}$$

where the $h_i^w(T)$ are weighting coefficients for different products.

Maximization of the final stock of goods

$$J = \sum_{i=1}^{n} h_i^x(T)x_i(T) \tag{3.17}$$

where the $h_i^x(T)$ are weighting coefficients ("costs") for $x_i(T)$.

Maximization of the terminal values of production capacity

$$J = \sum_{i=1}^{n} h_i^y(T)y_i(T) \tag{3.18}$$

where the $h_i^y(T)$ are weighting coefficients for $y_i(T)$.

Minimization of total expenses. This criterion is similar to the objective functions considered in Sections 1 and 2

$$J = \sum_{t=0}^{T-1} \beta(t)[(c^u(t), u(t)) + (c^v(t), v(t)) + (c^y(t), y(t))] \tag{3.19}$$

where

$c^u(t), c^y(t)$ are, respectively, operating and maintenance costs;
$c^v(t)$ is the investment cost; and
$\beta(t)$ is the discounting factor.

For storable goods [see inequalities (3.10)] it is desirable in some cases to extend eqn. (3.19) by including storage costs.

Other objective functions are of course also possible (Kantorovich 1965; V.V. Makarov 1966; Ivanilov and Petrov 1970a, b; Zimin 1976a, b, 1977, 1980). In addition, it should be noted that control targets can also be expressed by additional constraints, such as

$$\omega(T) \geqslant \bar{\omega}(T) \tag{3.20}$$

$$x(T) \geqslant \bar{x}(T) \tag{3.21}$$

$$y(T) \geqslant \bar{y}(T) \tag{3.22}$$

For example, one may wish to minimize the total costs [eqn. (3.19)] under a given level of final consumption as specified by inequality (3.20).

3.1.4 Statement of the Problem

For reference purposes we will now write down a typical optimization problem that frequently occurs in economic models.

Problem 3.1. Given the state equations of the production subsystem

$$x(t+1) = x(t) + [\mathrm{I} - \mathrm{A}(t)]\, u(t) - \sum_{\tau=0}^{\bar{\tau}} \mathrm{B}(\tau)\, v(t-\tau) - w(t) - s(t) \tag{3.1a}$$

and of the production-capacity subsystem

$$y(t+1) = [\mathrm{I} - \mathrm{D}(t)]\, y(t) + v(t-\bar{\tau}) \tag{3.4a}$$

with initial conditions

$$x(0) = x^0 \tag{3.2a}$$

$$v(t-\bar{\tau}) = v^0(t-\bar{\tau}) \qquad (0 \leqslant t \leqslant \bar{\tau} - 1) \tag{3.3a}$$

$$y(0) = y^0 \tag{3.5a}$$

find controls $u = \{u(0), \ldots, u(T-1)\}$, $v = \{v(0), \ldots, v(T-\bar{\tau}-1)\}$, and $w = \{w(0), \ldots, w(T-1)\}$, and corresponding trajectories $x = \{x(0), \ldots, x(T)\}$ and $y = \{y(0), \ldots, y(T)\}$, which satisfy the following constraints

(a) nonnegativity

$$u(t) \geqslant 0;\ v(t) \geqslant 0;\ w(t) \geqslant 0;\ x(t) \geqslant 0;\ y(t) \geqslant 0 \tag{3.6a}$$

(b) labor availability

$$\mathrm{L}(t)u(t) \leqslant l(t) \tag{3.7a}$$

(c) resource availability

$$\mathrm{R}(t)u(t) \leqslant r(t) \tag{3.8a}$$

(d) production capacity

$$u(t) \leqslant y(t) \tag{3.9a}$$

(e) storable goods inventory

$$x(t) \leqslant \bar{x}(t) \tag{3.10a}$$

(f) nonstorable goods inventory

$$[\mathrm{I} - \mathrm{A}(t)]\ u(t) \geqslant \sum_{\tau=0}^{\bar{T}} \mathrm{B}(\tau)v(t-\tau) + w(t) + s(t) \tag{3.13a}$$

(g) consumption

$$w(t) \geqslant g(t)\omega(t) \tag{3.14a}$$

and maximize the objective function

$$J = \sum_{t=0}^{T-1} \beta(t)\omega(t) \tag{3.15}$$

3.2 Discussion

We will now consider some modifications and extensions of Problem 3.1.

3.2.1 Conversion Model

In many practical cases it is necessary to take into account the process of reconstruction (or conversion) of productive capacity (Ivanilov and Petrov 1970a, b). In this case three of the conditions given above should be replaced.

State equation (3.1) should be replaced by

$$x_i(t+1) = x_i(t) + u_i(t) - \sum_{j=1}^{n} a_{ij}(t)u_j(t) - \sum_{j=1}^{n} \sum_{\tau=0}^{T_j} b_{ij}(\tau)v_j(t-\tau)$$
$$- \sum_{j,s=1}^{n} \sum_{\tau=0}^{T_j^s} b_{ij}^s(\tau)v_j^s(t-\tau) - w_i(t) - s_i(t) \tag{3.23}$$

Here

- $v_j^s(t)$ is the additional productive capacity in sector j obtained from conversion of some sector-s capacity started during period t;
- $b_{ij}^s(t)$ are the capital coefficients of the conversion $s \to j$; and
- τ_j^s is the number of steps required for the conversion $s \to j$.

The state equation (3.4) is replaced by

$$y_i(t+1) = [1 - d_i(t)]\, y_i(t) + v_i(t - \tau_i) - \sum_{s=1}^{n} \sum_{\tau=0}^{\tau_i^s - 1} v_i^s(t - \tau) + \sum_{s=1}^{n} k_i^s(\tau_i^s) v_i^s(t - \tau_i^s) \quad (3.24)$$

where $k_i^s(t)$ is the conversion coefficient, which shows the increase in the productive capacity in sector i per unit of conversion activity $s \to i$.

3.2.2 Capital Stock Subsystem

In some cases it is more convenient to describe the development of the production subsystem in terms of capital stock rather than in terms of productive capacity. In these cases, instead of state equations (3.4) or (3.24) we must introduce state equations

$$c_i(t+1) = [1 - \tilde{d}_i(t)]\, c_i(t) + v_i(t - \tau_i) - \sum_{s=1}^{n} \sum_{\tau=0}^{\tau_i^s - 1} v_s^i(t - \tau) + \sum_{s=1}^{n} v_i^s(t - \tau_i^s) \quad (3.25)$$

where

- $c_i(t)$ is the capital stock in sector i during period t; and
- $\tilde{d}_i(t)$ is the depreciation factor.

In addition, the production capacity constraints (3.9) are replaced by

$$\gamma_i(t) u_i(t) \leqslant c_i(t) \qquad (i = 1, 2, \ldots, n) \quad (3.26)$$

where $\gamma_i(t)$ is the capital–output ratio. Finally, if no conversion activities are taking place in the system, then the last term on the right-hand side of eqn. (3.25) should be omitted.

3.2.3 Simplified Model

We will now describe a simplified version of Problem 3.1, which may be of interest for more long-range planning and more aggregated systems, such as the case of linking energy and economy submodels. To simplify the model we assume that the period is such that time lags can be ignored and we rule out the possibility of building up stocks of goods; furthermore, we do not consider conversion or reconstruction processes. With these assumptions, the problem can be formulated as follows.

Problem 3.1a. Given the state equations for the capital stock subsystem in the form

$$c(t+1) = [\mathbf{I} - \mathbf{D}(t)]\, c(t) + v(t)$$

with an initial state

$$c(0) = c^0$$

and subject to the following constraints

(a) balance equations

$$[\mathbf{I} - \mathbf{A}(t)]\,u(t) = \mathbf{B}(t)v(t) + w(t) + s(t)$$

(b) resource availability

$$\mathbf{L}(t)u(t) \leqslant l(t)$$

$$\mathbf{R}(t)u(t) \leqslant r(t)$$

(c) production

$$\Gamma u(t) \leqslant c(t)$$

(d) consumption

$$w(t) \geqslant g(t)\omega(t)$$

find controls $\{v(t),\ u(t),\ \omega(t)\}$, and a corresponding trajectory $\{c(t)\}$, which maximize the objective function

$$J = \sum_{t=0}^{T-1} \beta(t)\omega(t)$$

3.2.4 INTERLINK Model

The INTERLINK model was developed at the International Institute for Applied Systems Analysis (IIASA) by Zimin for modeling the economic development of a region (or country) in the IIASA system of energy development models. It represents a version of the dynamic multisector π-model (Ivanilov and Petrov 1970a, b); its structure is close to that outlined in Problem 3.1 and it is described in detail elsewhere (Zimin 1976a, b, 1977, 1980).

The typical dimensions of the INTERLINK model are as follows: there are 17 state equations (representing sectors of the economy) and 41 constraints for each period. Each period is five years long and there are ten such periods, giving a total planning horizon of 50 years. The corresponding linear programming problem has approximately 600 rows and 600 columns.

4 LINKING THE MODELS

In earlier sections of this report we considered three different models – of the energy supply system, of the primary resources system, and of the economic development system; the most important features of each model were formally presented in Sections 1, 2, and 3, respectively. Each of these models can be used individually for the assessment of energy, resources, and the development of various technologies.

However, this approach of separate, "piecemeal" analysis is limited in its possibilities because many important features of the systems which derive from their interactions with one another are missing. To overcome these deficiencies we need to build models of the whole interacting energy–resources–economy system; we must therefore investigate ways of linking individual models into a coherent whole. This new stage of energy-policy modeling has started relatively recently (A.A. Makarov and Melentjev 1973; Dantzig 1975a; Dantzig and Parikh 1975; Belyaev et al. 1976; Behling et al. 1977; Häfele and A.A. Makarov 1977; Hitch 1977; Hoffman and Jorgenson 1977; Kononov 1977; A.A. Makarov 1977; Manne 1977). Two basic approaches* can be singled out here. In the first approach separate models are integrated into a single optimization problem with one corresponding objective function (Dantzig 1975a, b; Dantzig and Parikh 1975; Dantzig 1976). The second approach is to investigate manually linking a number of independent submodels, each with its own objective function (Behling et al. 1977; Häfele and A.A. Makarov 1977; Hoffman and Jorgenson 1977; A.A. Makarov 1977; Manne 1977).

Both approaches naturally have their own advantages and drawbacks. The major advantage of the first, "machine" approach is that it allows us to take into account all the constraints and interactions between the many factors which influence a given decision and to combine them in some "optimal" way. However, building an integrated model obviously leads to a very large optimization problem which, although sometimes possible to solve, is always very difficult to interpret.

The second, "manual" approach – in which information obtained from one submodel is interpreted by an analyst and provided as input to another submodel – is more attractive but is much more time consuming and may sometimes lead to uncertainty as to whether the "truly optimal" solution for the whole system has been obtained. Later in the report we will refer to this as the "iterative" approach.

It seems sensible to combine the best features of each approach and we will now consider each in turn, starting with the integrated model.

4.1 Integrated Model

Considering the ESS and the economy models, we can see (Figure 2) that there are two main links between them: the final demand for energy, which is an output of the economy model, and the demands for nonenergy resources, which are outputs of the ESS model. We will combine the ESS model (Problem 1.1) and the economy model (Problem 3.1) into one overall system, using the subscripts E for the energy sector and NE for the nonenergy sectors.

For uniformity of presentation we assume that the industrial processes of both the economic and the energy sectors may be described in terms of physical flows. Furthermore, in the model developed below we omit, for simplicity, time lags in the construction and putting into operation of production capacity; in other words, we will use simplified versions of the ESS and economy models.

---

* "Non-optimization" approaches fall outside the scope of this report and are therefore not considered here (see Hitch 1977).

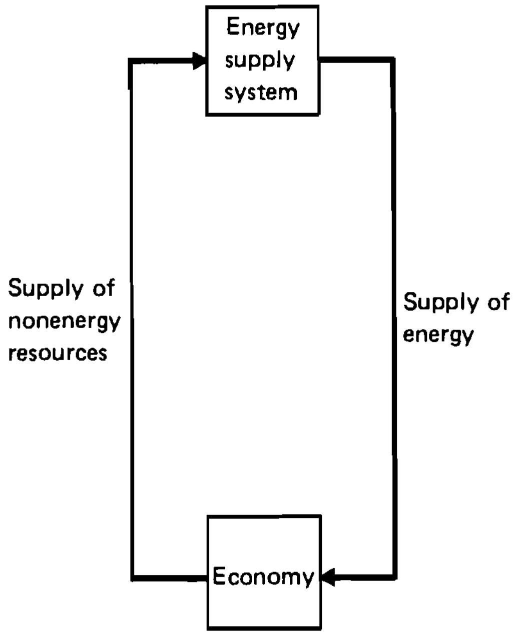


FIGURE 2 Linkage of energy supply and economy models.

4.1.1 State Equations

Production subsystem. This is a combination of state equations (1.1a) and (3.4a), for the energy and nonenergy sectors, respectively, in their simplified form (we describe the depreciation of capacity in the same way for both equations)

$$y_E(t+1) = [\mathbf{I} - \Delta_E(t)]y_E(t) + \nu_E(t) \tag{4.1}$$

$$y_{NE}(t+1) = [\mathbf{I} - \Delta_{NE}(t)]y_{NE}(t) + \nu_{NE}(t) \tag{4.2}$$

with initial states

$$y_E(0) = y_E^0 \tag{4.3}$$

$$y_{NE}(0) = y_{NE}^0 \tag{4.4}$$

Here $y_E(t)$ and $y_{NE}(t)$ are vectors of production capacity for the energy and nonenergy sectors, and $\nu_E(t)$ and $\nu_{NE}(t)$ are the increases of capacity in these sectors during period t. $\Delta_E(t)$ and $\Delta_{NE}(t)$ are diagonal matrices whose elements are the corresponding depreciation factors.

Energy resource consumption subsystem. To describe the cumulative consumption of primary energy resources we will first use eqn. (1.5a) (instead of the more-detailed version given in Problem 2.1)

$$z_E(t+1) = z_E(t) + \mathbf{Q}_E(t)u_E(t) \quad (4.5)$$

$$z_E(0) = z_E^0 \quad (4.6)$$

$$0 \leq z_E(t) \leq \bar{z}_E(t) \quad (4.7)$$

Here

$z_E(t)$ is the vector of cumulative amounts of primary energy resources extracted at the beginning of period t

$u_E(t)$ is the vector of activities in the energy sector.

The upper limits $\bar{z}_E(t)$ may be estimated from the resource model (see Section 2).

4.1.2 Constraints

The most important constraint in the model is the balance between the production of goods and their consumption. As in the simplified version of the economy model (Problem 3a), we rule out the possibility of building up stocks of goods, and therefore consider the static form of these conditions. For energy output

$$-\mathbf{A}_{NE}^{E}(t)u_{NE}(t) + [\mathbf{I} - \mathbf{A}_E^E(t)]u_E(t) = \mathbf{B}_{NE}^{E}(t)v_{NE}(t) + \mathbf{B}_E^E(t)v_E(t) + w_E(t) + s_E(t) \quad (4.8)$$

and for nonenergy products

$$[\mathbf{I} - \mathbf{A}_{NE}^{NE}(t)]u_{NE}(t) - \mathbf{A}_E^{NE}u_E(t) = \mathbf{B}_{NE}^{NE}(t)v_{NE}(t) + \mathbf{B}_E^{NE}(t)v_E(t) + w_{NE}(t) + s_{NE}(t) \quad (4.9)$$

We also have production-capacity constraints for energy sectors

$$u_E(t) \leq y_E(t) \quad (4.10)$$

and for nonenergy sectors

$$u_{NE}(t) \leq y_{NE}(t) \quad (4.11)$$

[essentially similar in form to inequalities (1.4) and (3.9), respectively].

Labor-availability constraints (3.7) are written in the form

$$\mathbf{L}_{NE}(t)u_{NE}(t) + \mathbf{L}_E(t)u_E(t) \leq l(t) \quad (4.12)$$

and the constraints on WELMM factors [cf. inequality (3.8)] as

$$\mathbf{R}_{NE}(t)u_{NE}(t) + \mathbf{R}_E(t)u_E(t) \leq r(t) \quad (4.13)$$

Final consumption constraints (3.14) can be written as

$$w_E(t) \geqslant g_E(t)\omega(t) \tag{4.14}$$

$$w_{NE}(t) \geqslant g_{NE}(t)\omega(t) \tag{4.15}$$

where the given vectors $g_{NE}(t)$ and $g_E(t)$ specify profiles of final consumption for non-energy and energy products, respectively.

Finally, all the variables are obviously nonnegative

$$\begin{gathered} u_{NE}(t) \geqslant 0;\ u_E(t) \geqslant 0;\ v_{NE}(t) \geqslant 0;\ v_E(t) \geqslant 0; \\ y_{NE}(t) \geqslant 0;\ y_E(t) \geqslant 0;\ z_E(t) \geqslant 0;\ \omega(t) \geqslant 0 \end{gathered} \tag{4.16}$$

4.1.3 Statement of the Problem

We therefore obtain the following optimization problem.

Problem 4.1. Given the state equations

$$y_E(t+1) = [\mathbf{I} - \Delta_E(t)]y_E(t) + v_E(t) \tag{4.1}$$

$$y_{NE}(t+1) = [\mathbf{I} - \Delta_{NE}(t)]y_{NE}(t) + v_{NE}(t) \tag{4.2}$$

with initial states

$$y_E(0) = y_E^0 \tag{4.3}$$

$$y_{NE}(0) = y_{NE}^0 \tag{4.4}$$

find controls $\{v_E(t)\}$, $\{v_{NE}(t)\}$ and $\{u_E(t)\}$, $\{u_{NE}(t)\}$, and corresponding trajectories $\{y_E(t), y_{NE}(t)\}$, which satisfy the following constraints

(a) balance equations

$$[\mathbf{I} - \mathbf{A}_E^E(t)]u_E(t) - \mathbf{A}_{NE}^E(t)u_{NE}(t) = \mathbf{B}_E^E(t)v_E(t) + \mathbf{B}_{NE}^E(t)v_{NE}(t) + w_E(t) + s_E(t) \tag{4.8}$$

$$\begin{aligned} -\mathbf{A}_E^{NE}(t)u_E(t) + [\mathbf{I} - \mathbf{A}_{NE}^{NE}(t)]u_{NE}(t) &= \mathbf{B}_E^{NE}(t)v_E(t) + \mathbf{B}_{NE}^{NE}(t)v_{NE}(t) \\ &\quad + w_{NE}(t) + s_{NE}(t) \end{aligned} \tag{4.9}$$

(b) production capacity

$$u_E(t) \leqslant y_E(t) \tag{4.10}$$

$$u_{NE}(t) \leqslant y_{NE}(t) \tag{4.11}$$

(c) primary energy resource availability

$$z_E(t+1) = z_E(t) + Q_E(t)u_E(t) \tag{4.5}$$

$$z_E(0) = z_E^0 \tag{4.6}$$

$$z_E(t) \leqslant \bar{z}(t) \tag{4.7}$$

(d) labor availability

$$L_E(t)u_E(t) + L_{NE}(t)u_{NE}(t) \leqslant l(t) \tag{4.12}$$

(e) WELMM factor availability

$$R_E(t)u_E(t) + R_{NE}(t)u_{NE}(t) \leqslant r(t) \tag{4.13}$$

(f) final consumption

$$w_E(t) \geqslant g_E(t)\,\omega(t) \tag{4.14}$$

$$w_{NE}(t) \geqslant g_{NE}(t)\,\omega(t) \tag{4.15}$$

(g) nonnegativity

$$\begin{aligned} &u_E(t) \geqslant 0;\ u_{NE}(t) \geqslant 0;\ v_E(t) \geqslant 0;\ v_{NE}(t) \geqslant 0; \\ &y_E(t) \geqslant 0;\ y_{NE}(t) \geqslant 0;\ z_E(t) \geqslant 0;\ \omega(t) \geqslant 0 \end{aligned} \tag{4.16}$$

and which maximize the objective function*

$$J = \sum_{t=0}^{T-1} \beta(t)\,\omega(t) \tag{4.17}$$

Problem 4.1 is, once again, a DLP model. Its solution , in principle, permits us to investigate the interactions between a (more-detailed) energy sector and the nonenergy sectors of an economy. As mentioned above, we can solve Problem 4.1 as one overall DLP problem, or we can solve it by an iterative procedure, paying special attention to the links between the ESS and the economy parts of the integrated model.

Clearly, in much the same way, the more-detailed statement of the resources model (Problem 2.1) may be included in the integrated model instead of using the simplified eqns. (4.5)–(4.7). We will not, however, develop this possibility here.

In the integrated model there is one important feature which, although clearly visible in the scalar representation, cannot be seen explicitly from the matrix formulation of Problem 4.1. In practise, each of the individual models which are to be integrated into a

* This particular objective function is chosen only for illustrative purposes. Many other objectives are of course of interest for this integrated model.

system may have different levels of aggregation. Moreover, if we are investigating the influence of ESS on economic development, the ESS model should be presented in much more detail than the economy model. For this particular case, a special model has been developed (see below) which determines the influence (or impact) of energy developments upon the economy as a whole.

Therefore, when attempting the linkage of energy, resources, and economy models, one must take into account first, the means of linkage (machine or man–machine), and second, the level of aggregation and specific features of each individual model.

4.2 Iterative Approach

We now consider the iterative interaction between ESS and economy model. The general scheme is as follows.

On examining the integrated model described earlier (Problem 4.1), we see that it is basically the economy model (Problem 3.1) partitioned into energy (E) and nonenergy (NE) sectors. On the other hand, the ESS model is embedded in the integrated model. In fact, eqns. (4.1), (4.3), (4.5)–(4.7), (4.11), and (4.14) are the same as in the Problem 1.1 formulation.

If we define the demand $d_E(t)$ for secondary energy by

$$d_E(t) = \mathbf{A}_{NE}^{E}(t)u_{NE}(t) + \mathbf{B}_{NE}^{E}(t)v_{NE}(t) + w_E(t) + s_E(t) \tag{4.18}$$

and let

$$\mathbf{D}_E(t) = [\mathbf{I} - \mathbf{A}_E^E(t)] \tag{4.19}$$

then we can rewrite eqn. (4.8) as

$$\mathbf{D}_E(t)u_E(t) = d_E(t) + \mathbf{B}_E^E(t)v_E(t) \tag{4.19}$$

which, because of the smallness of the last term on the right-hand side, is similar to the demand constraints (1.11) of the ESS model.

Let us further write down the requirements of the ESS for nonenergy products as follows

$$f_E^{NE}(t) = \mathbf{B}_E^{NE}(t)v_E(t) + \mathbf{A}_E^{NE}(t)u_E(t) \tag{4.20}$$

Taking into account that the amounts of nonenergy products required for the operation and maintenance of energy production systems [the second term on the right-hand side of eqn. (4.20)] are small in comparison with the requirements for construction [the first term on the right-hand side of eqn. (4.20)], it can be seen from eqn. (4.20) and inequality (1.9), that

$$\mathbf{F}(t) = \mathbf{B}_E^{NE}(t)$$

Therefore, we can rewrite eqn. (4.9) as

$$[\mathbf{I} - \mathbf{A}_{NE}^{NE}(t)]\, u_{NE}(t) = f_{NE}^{NE}(t) + f_{E}^{NE}(t) \tag{4.21}$$

where

$$f_{NE}^{NE}(t) = \mathbf{B}_{NE}^{NE}(t) v_{NE}(t) + w_{NE}(t) + s_{NE}(t) \tag{4.22}$$

and $f_{E}^{NE}(t)$ is defined from eqn. (4.20).

Thus, eqn. (4.19) represents the supply of energy required for the energy sector and, as was mentioned above, is equivalent to the demand constraint (1.11) with $d_E(t)$ fixed; and constraint (4.20) represents the amounts of nonenergy products required by the ESS for a fixed value of $f_{E}^{NE}(t)$.

On the other hand, eqns. (4.18) and (4.22), respectively, represent the demands for energy and nonenergy products in the rest of the economy, while eqn. (4.21) shows the supply of goods from the nonenergy sectors.

In addition, we can rewrite constraints (4.12) and (4.13) in the following form

$$\mathbf{L}_E(t) u_E(t) = l_E(t) \tag{4.23}$$

$$\mathbf{L}_{NE}(t) u_{NE}(t) = l_{NE}(t) \tag{4.24}$$

$$l_E(t) + l_{NE}(t) \leqslant l(t) \tag{4.25}$$

$$\mathbf{R}_E(t) u_E(t) = r_E(t) \tag{4.26}$$

$$\mathbf{R}_{NE}(t) u_{NE}(t) = r_{NE}(t) \tag{4.27}$$

$$r_E(t) + r_{NE}(t) \leqslant r(t) \tag{4.28}$$

Finally, we find that eqns. (4.1), (4.3), (4.5)–(4.7), (4.10), (4.14), (4.19), (4.20), (4.23), and (4.26), with variables $d_E(t)$, $f_{E}^{NE}(t)$, $l_E(t)$, and $r_{NE}(t)$ given exogenously, give a complete description of the ESS model; similarly, eqns. (4.2), (4.4), (4.11), (4.15), (4.18), (4.21), (4.22), (4.24), and (4.27), with variables $d_E(t)$, $f_{E}^{NE}(t)$, $l_{NE}(t)$, and $r_{NE}(t)$ given exogenously, describe the rest of the economy.

In the integrated model (Problem 4.1), variables $d_E(t)$, $f_{E}^{NE}(t)$, $f_{NE}^{NE}(t)$, $l_E(t)$, $l_{NE}(t)$, $r_E(t)$, and $r_{NE}(t)$, should be considered as endogenous; in this case constraints (4.21), (4.25), and (4.28) are coupling constraints and the variables just mentioned $[d_E(t)$, etc.] are coupling variables.

Let us assume that we have some initial estimate of the energy demand $\bar{d}_E(t)$ for a given planning period $0 \leqslant t \leqslant T-1$. Solving the ESS model (Problem 1.1) for this demand, we can calculate the required increases in capacity $\bar{v}_E(t)$ of the ESS during the period, and the corresponding values for the production capacity $\bar{y}_E(t)$ and output (degrees of utilization) $\bar{u}_E(t) \leqslant \bar{y}_E(t)$.

The requirements of the ESS in nonenergy resources, $\bar{f}_{E}^{NE}(t)$, are calculated from eqn. (4.21). Now we can solve the economy model (Problem 3.1) or the integrated model

(Problem 4.1) with fixed $\bar{u}_E(t)$, $\bar{v}_E(t)$, $\bar{y}_E(t)$, subject to a certain set of assumptions about the future development of the overall economy.

This solution yields degrees of utilization (gross outputs) $\bar{u}_{NE}(t)$ and the additional capital investments $\bar{v}_{NE}(t)$ required in the nonenergy sectors as well as a new value $\bar{d}^*_E(t)$ for the corresponding demand for energy [calculated from eqn. (4.18)]. If the old $\bar{d}_E(t)$ and new $\bar{d}^*_E(t)$ values for energy demand coincide, the procedure terminates; if the values do not coincide, then we must repeat the iteration with a recalculated demand.

Generally speaking, the solution obtained in such a way (if the process converges) is not an optimal solution for Problem 4.1, but is often acceptable because it satisfies all the constraints of the problem and optimizes (separately) two objectives [for example (1.12) and (3.15)] for the energy and nonenergy sectors.

To obtain an optimal solution for the whole of Problem 4.1 by an iterative procedure, one may use different methods of decomposition. In this case the dual variables (marginal estimates), obtained from the solution of the economy model, define the corresponding objective function for the ESS model [instead of using eqn. (1.12)]. The actual convergence behavior depends on the procedure used and the method of implementation. It should also be noted that for this procedure to be implemented the economy model should be sufficiently disaggregated in order to provide the ESS model with shadow prices in sufficient detail.

But, in practice, a single "optimal" solution of Problem 4.1 is not very valuable – regardless of whether it has been obtained "automatically" by applying the simplex method to Problem 4.1, or in some iterative way. Clearly, such a complex system requires a man–machine iterative procedure with a detailed energy–economy analysis composed of separate iterations. Let us now examine the points where human intervention is appropriate. These are as follows

- Changing the objective function for the overall Problem 4.1 and for the ESS model (Problem 1.1). [In fact, this is a vector-optimization problem (Alta Conference 1975)].
- Determining the energy demand $d_E(t)$ not from eqn. (4.18), but rather from a special energy-demand model (see for example Beaujean et al. 1977).
- Determining the nonenergy resource requirements $f_E^{NE}(t)$ for the ESS by using a special model (see Kononov and Tkachenko 1975).
- Changing the parameters of the model (especially those associated with assumptions on technological innovation and profiles of consumption).

Many of these points of human intervention may be considered as attempts to take into account nonlinearities of the system.

It should be noted finally that the methodological problems of linking separate models into coherent overall systems are of great practical importance and have not yet been sufficiently investigated. Some of these questions are discussed at greater length by Kallio et al. (1979).

4.3 Discussion

4.3.1 Pilot Model

This model (Dantzig 1975a, b; Dantzig and Parikh 1975; Dantzig 1976) has been developed by Dantzig and Parikh at Stanford University. It is a DLP model on a pilot scale that describes, in physical terms, various technological interactions within the sectors of the US economy, including a detailed energy sector.

The basic structure of the model is quite similar to that described by Problem 4.1. Dynamic equations include capacity-balance constraints, retraining of labor force constraints. and constraints on raw energy reserves, cumulative discoveries, amounts produced, and intermediate energy stocks.

The capacity-balance constraints are equivalent to eqns. (4.1) and (4.2). The retraining of labor force constraints specify educational and training capacities of the country modeled and are written in the form [compare inequalities (1.27) and (1.34) in the DESOM model described in Section 1.2.4]

$$p(t+1) \leqslant \beta p(t)$$

where the manpower vector $p(t)$ is partitioned into skill groups.

The resource constraints are similar to constraints (2.24) and (2.25) and are intended to allow the inclusion of accurate values for the energy reserves, cumulative discoveries (and amounts produced), and stocks.

The various static constraints represent energy-demand requirements, energy-processing and operating-capacity limitations, and environmental aspects. The energy and non-energy sectors are linked by the balance equation constraints (4.8) and (4.9).

The objective function of the model maximizes the discounted vector of goods received per person, summed over time. It can be expressed as

$$J = \sum_{t=1}^{T} \lambda(t)[\mathbf{M}(t), p(t)]$$

where the matrix $\mathbf{M}(t)$ represents the consumption levels and the vector $p(t)$ is the distribution of the population over different income levels.

When finally completed, the detailed model will include an 87-sector input–output matrix, and the possibility of modeling the energy sector using approximately 150 equations per period. Thus, the number of constraints for each period in an integrated model with a reasonable level of detail may be of the order of 400: 87 for industrial activity, 2 X 87 for capacity constraints, and about 150 for a detailed energy sector. A 20–25-period model (for example, one covering a 75-year planning horizon in 3-year periods) would therefore have between 8,000 and 10,000 constraints.

As noted by Dantzig (1976), such LP models would be among the largest built to date. Therefore as a first step, a much smaller model which (Dantzig 1976) "incorporates many, if not all, of the essential features of its larger counterpart" has been attempted. This pilot model is expected to have about 130 equations per period. For a 30-year model (ten periods of three years each), there will be between 1,250 and 1,400 equations. Initially, the model will be solved using the straightforward simplex method.

4.3.2 IMPACT Model

This is an extension of the model developed by Kononov and Tkachenko at the Siberian Power Institute (Kononov and Tkachenko 1975; Kononov and Por 1979). The model is designed to investigate the influence upon other branches of the national economy of long-term changes in technology and the structure and rate of energy development.

The model is described by the following equations [for more details, see Kononov and Por (1979)].

The direct requirements of the ESS for nonenergy products are given by

$$f_{E}^{NE}(t) = A_{E}^{NE}(t)u_{E}(t) + \sum_{\tau=0}^{\bar{\tau}} B_{E}^{NE}(t-\tau)v_{E}(t-\tau) \tag{4.29}$$

If we neglect the time lags $\bar{\tau}$ in construction, then eqn. (4.20) is obtained. In the original version of the IMPACT model (Kononov and Por 1979), a "carried forward" presentation is used; in other words

$$f_{E}^{NE}(t) = A_{E}^{NE}(t)u_{E}(t) + \sum_{\tau=t}^{t+\bar{\tau}} \tilde{B}_{E}^{NE}(\tau-t)v_{E}(\tau) \tag{4.29a}$$

where the matrix $\tilde{B}_{E}^{NE}(\tau - t)$ denotes the contribution for the construction of additional capacity to be put into operation during period $\bar{\tau}$, where $t \leqslant \tau \leqslant t + \bar{\tau}$.

Total (direct and indirect) product (material, equipment, etc.) requirements are derived from eqn. (4.9) [or from eqn. (4.21), where $f_{E}^{NE}(t)$ and $f_{NE}^{NE}(t)$ are obtained from eqns. (4.22) and (4.29), respectively]

$$[I - A_{NE}^{NE}(t)]u_{NE}(t) = B_{NE}^{NE}(t)v_{NE}(t) + f_{E}^{NE}(t) + w_{NE}(t) + s_{NE}(t) \tag{4.30}$$

Using $v_{NE}(t)$ and $v_{E}(t)$, one can also calculate the total direct and indirect capital investments. In addition, the model includes several equations for evaluating direct and indirect expenditures of WELMM resources.

The model operates in the following way. Problem 1.1 for the given demand $\bar{d}_{E}(t)$ for secondary energy is solved. Initially, the nonenergy resource constraints (1.9) are not taken into account. The solution of the problem gives the values $\bar{u}_{E}(t)$ and $\bar{v}_{E}(t)$, which are inputs for the IMPACT model. Using eqn. (4.29), one can calculate $\bar{f}_{E}^{NE}(t)$ for given $\bar{u}_{E}(t)$ and $\bar{v}_{E}(t)$. Substituting $\bar{f}_{E}^{NE}(t)$ into eqn. (4.30) and solving the linear equations (4.24) with certain additional conditions (Kononov and Por 1979)*

$$v_{NE}(t) = \max\{\min_{\tau<t} [u_{NE}(t) - u_{NE}(\tau)]\,;\ 0\}$$

one can find the indirect investment $v_{NE}(t)$ in the economy which the ESS needs to meet the given demand $\bar{d}_{E}(t)$.

Note that we have only described here the general scheme of the IMPACT model. The particular implementation of this model depends greatly on the specific details of the ESS and economy models to be linked.

---

* It is assumed here that capital stock is not dismantled and does not wear out.

4.3.3 SPI Model

The interactions between the energy and nonenergy sectors of the national economy have also been analyzed at the Siberian Power Institute, part of the Siberian Branch of the USSR Academy of Sciences. For this analysis a special multisector model has been developed (A.A. Makarov 1977). The model describes the interactions of the energy (E) sector with those nonenergy (NE) sectors which directly or indirectly influence the energy sector. There are eight such nonenergy sectors producing a total of 31 types of product.

The mathematical formulation of the model is close to that described by Problem 4.1 [note that we use here a somewhat different notation from that in the original version of the model (A.A. Makarov 1977)].

The development of the production subsystem is described by state equations which are similar to eqns. (4.1) and (4.2)

$$\sum_e y_{ie\mathrm{E}}(t+1) = \sum_e \delta_i(t) y_{ie\mathrm{E}}(t) + \sum_e v_{ie\mathrm{E}}(t)$$

$$y_{i\mathrm{NE}}(t+1) = \delta_i(t) y_{i\mathrm{NE}}(t) + v_{i\mathrm{NE}}(t)$$

where $\delta_i(t)$ is a depreciation factor. Note that the subscript e used in equations for the energy sector denotes e different technologies for energy production. Thus the energy sector is represented in a more disaggregated form in comparison to the nonenergy sectors of the model.

The balance equations are written in the dynamic form [compare eqns. (3.1), (4.8), and (4.9)]

for the nonenergy sectors

$$z_{i\mathrm{NE}}(t+1) = z_{i\mathrm{NE}}(t) + a_{i\mathrm{NE}}(t) y_{i\mathrm{NE}}(t) - \sum_j a_{ij\mathrm{NE}}(t) y_{j\mathrm{NE}}(t) - \sum_j \sum_\tau b_{ij\mathrm{NE}}(t+\tau) v_{j\mathrm{NE}}(t+\tau) - w_{i\mathrm{NE}}(t) - s_{i\mathrm{NE}}(t)$$

for the energy sector

$$z_{i\mathrm{E}}(t+1) = z_{i\mathrm{E}}(t) + \sum_e a_{ie\mathrm{E}}(t) y_{ie\mathrm{E}}(t) - \sum_j a_{ij\mathrm{E}}(t) y_{j\mathrm{NE}}(t) - w_{i\mathrm{E}}(t) - s_{i\mathrm{E}}(t)$$

(For the energy sector the stocks are fuels.)

Here $z_{i\mathrm{NE}}(t)$ and $z_{i\mathrm{E}}(t)$ are the production inventories for the nonenergy and energy sectors, respectively, at the beginning of period t; $a_{i\mathrm{NE}}(t)$ and $a_{ie\mathrm{E}}(t)$ are loading coefficients of production capacity, hence

$$u_{i\mathrm{NE}}(t) = a_{i\mathrm{NE}}(t) y_{i\mathrm{NE}}(t)$$

$$u_{ie\mathrm{E}}(t) = a_{ie\mathrm{E}}(t) y_{ie\mathrm{E}}(t)$$

where $u_{i\mathrm{NE}}(t)$ and $u_{ie\mathrm{E}}(t)$ are the production levels (gross outputs) during period t.

As in the IMPACT model, a "carried forward" ($\tau > 0$) presentation of the requirements for construction is used.

Constraints on manpower and other limited resources are given in a similar way to inequalities (4.12) and (4.13)

$$\sum_i \sum_e r^{\mathrm{E}}_{vie}(t) y_{ie\mathrm{E}}(t) + \sum_e \sum_\tau r^{\mathrm{E}}_{vie}(t,\tau) v_{ie\mathrm{E}}(\tau) + r^{\mathrm{NE}}_{vi}(t) y_{i\mathrm{NE}}(t)$$

$$+ \sum_\tau r^{\mathrm{NE}}_{vi}(t,\tau) v_{i\mathrm{NE}}(\tau) \leqslant r_v(t)$$

The model is solved using an iterative mode.

5 DLP CANONICAL FORM

On considering the models described above, we can see that all of them can be reduced to a single canonical form (Propoi 1973, 1976).

Problem 5.1. Given the state equations

$$x(t+1) = \mathrm{A}(t)x(t) + \sum_{\tau=0}^{\bar{\tau}} \mathrm{B}(\tau)u(t-\tau) \tag{5.1}$$

with initial conditions

$$x(0) = x^0;\ u(t-\tau) = u^0(t-\tau) \qquad (0 \leqslant \tau \leqslant t-1) \tag{5.2}$$

and constraints

$$\mathrm{G}(t)x(t) + \mathrm{D}(t)u(t) \leqslant f(t) \tag{5.3}$$

$$x(t) \geqslant 0;\ u(t) \geqslant 0 \tag{5.4}$$

find the control $u = \{u(0), \ldots, u(T-1-\tau)\}$ and the corresponding trajectory $x = \{x(0), \ldots, x(T)\}$, which maximize the objective function

$$J(u) = [a(T), x(T)] + \sum_{t=0}^{T-1} [(a(t), x(t)) + (b(t), u(t))] \tag{5.5}$$

Here the $\{u(t)\}$ are control variables and the $\{x(t)\}$ are state variables.

One can see that either all the models considered in the previous sections can be reduced to this canonical DLP problem, or that the methods developed for the canonical problem can be directly applied to the models. Problem 5.1 represents a DLP problem in a canonical form and can be viewed either as a "staircase" linear-programming problem or as an optimal control-theory problem. Hence, both methods – linear programming and control theory – can be applied to the solution of Problem 5.1. These methods have been surveyed by Propoi (1973, 1976, 1979).

6 CONCLUSION

Different individual energy–resource–economy models, and their linkage into a coherent overall system, have been discussed in the preceding sections. It has been shown that all these models may be reduced to a canonical form of the DLP problem. Therefore, a unified methodological approach can be developed to analyze and solve the models. Very briefly, several further possible directions for the methodological analysis of energy models may be outlined.

a. Energo-economic analysis. In this report we have concentrated on analyzing the common mathematical features of the models. The analysis of the physical structure of each model – objective functions, constraints, level of aggregation, uniformity of data bank, etc. – from the economic and energy-technology points of view is also of great interest.

b. Vector-optimization methods. Clearly, a single objective function is not a realistic way of modeling energy systems. This problem has been discussed, for example, by Ho (1979).

c. Duality theory. The shadow prices which are the solutions of the dual problem provide a valuable tool for a marginal analysis of the model. The relevant duality theory for the canonical DLP Problem 5.1 has been described by Propoi (1977). The further application of this theory to energy models, as discussed in this report would be useful in many respects.

d. Numerical-solution methods. As mentioned above, Problem 5.1 is an LP problem. Hence, standard LP programs can be (and already have been) applied for the solution of energy models. Special methods which take into account the specific features of DLP problems have also been developed (Ho and Manne 1974; Propoi and Yadykin 1975/1976; Ho 1977; Ho and Loute 1977; Propoi and Krivonozhko 1977, 1978); see also the references given by Propoi (1976, 1979). Preliminary versions of these algorithms show results which are acceptable when compared to the standard simplex methods (Ho 1977; Ho and Loute 1977).

e. Post-optimal analysis. Methods for analysis of solutions, including parametric DLP methods, and sensitivity and stability analysis, are of great practical interest. A general theory of linear and quadratic parametric programming has recently been developed (Propoi and Yadykin 1978).

f. Implementation of the solution. The implementation of the optimal solution is just as important as finding the solution. We must mention here the questions of realization of the optimal solution as a program (that is, as a time sequence of controlling actions) or as a feedback control (that is, as a current control action determined by the current state of the system).

g. Linking the models. The development of methods for linking individual models is, at present, probably the most important issue. Three main areas for investigation are

- Relations between long-, medium-, and short-term energy models (for example, how the optimal solution of an aggregated long-term model relates to the solution of a more detailed short-term model);
- Methods of linking individual models of energy, resources, and the economy into an integrated energy model for a nation or region (some of these methodological questions have been discussed in Section 4 of this report); and
- methods of linking national energy models into a world model.

Various discussions, both of methodology and of actual methods for the computer implementation of linked models, may be found in the literature (see, for example, Moiseev 1975; Behling et al. 1977; Häfele and A.A. Makarov 1977; Hoffman and Jorgenson 1977; Kononov 1977; A.A. Makarov 1977; Manne 1977; Moiseev 1977; Orchard-Hays 1977; Kallio et al. 1979).

ACKNOWLEDGMENTS

During the preparation of this report, we have had many helpful discussions with colleagues and received numerous suggestions on the subject. In particular we should mention the very detailed and valuable comments of Prof. A.A. Makarov, Prof. Yves Smeers, and the anonymous reviewer. These comments and discussions helped us enormously and we are sincerely grateful for them. However, we of course accept full responsibility for any misinterpretations or inaccuracies which may still be present in the text.

REFERENCES

Aganbegyan, A.G., and K.K. Valtukh (1975) Ispolzovanie Narodno Khozyajstvennykh Modelej v Planirovanii (Use of National-Level Economic Models in Planning). Moscow: Ekonomika (in Russian).

Aganbegyan, A.G., L.A. Kozlov, and D.M. Kazakevich (Eds.) (1974) Planirovanie Otraslevykh Sistem: Modeli i Metody Optimizatsii (Planning of Multisectoral Systems: Models and Methods of Optimization). Moscow: Ekonomika (in Russian).

Agnew, M., L. Schrattenholzer, and A. Voss (1978a) A Model for Energy Supply Systems and Their General Environmental Impact. RM-78-26. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Agnew, M., L. Schrattenholzer, and A. Voss (1978b) User's Guide for the MESSAGE Computer Program. RM-78-26, Addendum. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Alta Conference (1975) Energy: Mathematics and Models. Proceedings of a Conference held at Alta, Utah, July.

Ayres, R.U. (1978) Resources, Environment, and Economics: Applications of the Materials/Energy Balance Principle. New York: John Wiley.

Beaujean, J.-M., J.P. Charpentier, and T. Müller (1977) Energy Demand Studies: Modelling Useful Energy. Discussion Paper, IIASA Workshop on Energy Strategies, May 1977. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Behling, D.J., Cherniavsky, E.A., and K.C. Hoffman (1977) Overview of Technological–Economic Models for Energy Policy Analysis. IIASA Workshop on Energy Strategies, May 1977. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Belyaev, L.S., Yu.D. Kononov, and A.A Makarov (1976) Methods and Models for the Optimization of Energy Systems Development: Soviet Experience. In J.-M. Beaujean and J.-P. Charpentier (Eds.), A Review of Energy Models, No. 3. RR-76-18. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Brobst, D.A., and W.P. Pratt (1973) United States Mineral Resources. Washington, D.C.: United States Geological Survey.

Dantzig, G.B. (1975a) Formulating a Pilot Model for Energy in Relation to the National Economy. Report SOL 75-10. Stanford, California: Stanford University Systems Optimization Laboratory.

Dantzig, G.B. (1975b) A Note on the Objective Function for the Pilot Model. Report SOL 75-20. Stanford, California: Stanford University Systems Optimization Laboratory.

Dantzig, G.B. (1976) A Pilot Linear Programming Model for Assessing Physical Impact on the Economy of a Changing Energy Picture. IIASA Conference, May 1976. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Dantzig, G.B., and S.C. Parikh (1975) On a Pilot Linear Programming Model for Assessing Physical Impact on the Economy of a Changing Energy Picture. Report SOL 75-14R. Stanford, California: Stanford University Systems Optimization Laboratory.

Ermoliev, Yu. M. (1978) Stochastic Programming. Moscow: Nauka (in Russian).

Govett, G.J.S., and M.H. Govett (1974) The Concept and Measurement of Mineral Reserves and Resources. Resources Policy, September.

Grenon, M. (Ed.) (1976) Proceedings of the First IIASA Conference on Energy Resources, May 1975. CP-76-4. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Grenon, M., and B. Lapillone (1976) The WELMM Approach to Energy Strategies and Options. RR-76-19. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Grenon, M., and I. Zimin (1977) Resources Model. IIASA Workshop on Energy Strategies, May 1977. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Häfele, W. (1974) Future Energy Resources. RR-74-20. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Häfele, W., and A.A. Makarov (1977) Modeling of Medium- and Long-Range Energy Strategies. Discussion paper, IIASA Workshop on Energy Strategies, May 1977. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Häfele, W., and A.S. Manne (1974) Strategies for a Transition from Fossil to Nuclear Fuels. RR-74-7. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Häfele, W., and W. Sassin (1976) Energy Strategies. RR-76-8. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Häfele, W., M. Grenon, I. Zimin, B. Lapillone, C.R. Bell, W. Sassin, and H. Otway (1977) Report to the Advisory Committee of IIASA's Energy Program. RM-77-42. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Hitch, C.J. (Ed.) (1977) Modelling Energy-Economy Interactions: Five Approaches. Washington, D.C.: Resources for the Future.

Ho, J.K. (1977) Nested Decomposition of a Dynamic Energy Model. Management Science 23:9.

Ho, J.K. (1979) Hollistic Preference Evaluation in Multiple Criteria Optimization. BNL-25656, AMD-818. Upton, N.Y.: Brookhaven National Laboratory.

Ho, J.K., and E. Loute (1977) A Comparative Study of Two Methods for Staircase Linear Programs. C.O.R.E. Université Catholique de Louvain, Belgium.

Ho, J.K., and A. Manne (1974) Nested Decomposition for Dynamic Models. Mathematical Programming 6:2.

Hoffman, K.C., and D.W. Jorgenson (1977) Economic and Technological Models for Evaluation of Energy Policy. The Bell Journal of Economics, Autumn Issue.

Hudson, E.A., and D.W. Jorgenson (1975) In J.-P. Charpentier (Ed.), A Review of Energy Models, No. 2. RR-75-35. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Ivanilov, Yu.P., and A.A. Petrov (1970a) Dynamic Multisector Production Model (II-Model). Kybernetika 2 (in Russian).

Ivanilov, Yu.P., and A.A. Petrov (1970b) The Problem of Optimization in Dynamic Multi-Branch Industrial Models. Lecture Notes in Mathematics, Volume 112, Colloquium on Methods of Optimization. Berlin: Springer.

Kallio, M., W. Orchard-Hays, and A. Propoi (1979) Linkage of Optimization Models. Paper presented at the Tenth International Symposium on Mathematical Programming, Montreal, August 1979.

Kantorovich, L.V. (1965) The Best Use of Economic Resources. Cambridge, Massachusetts: Harvard University Press.

Kaya, Y., and Y. Suzuki (1974) Global Constraints and a New Vision for Development – I. Technological Forecasting and Social Change 6: 277–297.

Kononov, Yu.D. (Ed.) (1977) Methods of Systems Analysis for Long-Term Energy Development. CP-77-2. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Kononov, Yu.D., and A. Por (1978) The Economic IMPACT Model. RR-79-80. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Kononov, Yu.D., and V.Z. Tkachenko (1975) Method for Approximate Estimation of the Fuel-Energy-Complex Development Impact on Related Sectors of an Economy. In, Voprosy Vlijania Razvitia Energetiki na Drugie Otrasli Narodnovo Hoziajstva. Irkutsk: Siberian Branch of the USSR Academy of Sciences.

Krivonozhko, V.E., and A.I. Propoi (1979) On a Method for Evolution of Dynamic Transportation Problems. Automatika i Telemekhanika 12: 133–145.

Kydes, A.S. (1978) Brookhaven Energy System Optimization Model: its Variants and Uses. BNL 50873, UC-13. Upton, N.Y.: Brookhaven National Laboratory.

Makarov, A.A. (Ed.) (1977) Methods and Models for Analysis of Optimal Directions of Long-Range Development of the Energy System. Irkutsk: Siberian Branch of the USSR Academy of Sciences (in Russian).

Makarov, A.A., and L.A. Melentjev (1973) Metody Issledovania i Opimizatzii Energeticheskogo Khozi-astva (Analysis and Optimization Methods for Energy Systems). Novosibirsk: Nauka (in Russian).

Makarov, V.V. (1966) Linear Dynamic Models of Production. In: Optimalnoe Planirovanie V. Novosibirsk: Nauka (in Russian).

Manne, A.S. (1976) ETA: A Model for Energy Technology Assessment. The Bell Journal of Economics, Autumn Issue.

Manne, A.S. (1977) ETA-MACRO: A Model of Energy–Economy Interactions. ORSA/TIMS Meeting, San Francisco, May 1977.

Marcuse, W., et al. (1976) A Dynamic Time-Dependent Model for the Analysis of Alternative Energy Policies. In K.B. Haley (Ed.), Operational Research '75. Amsterdam: North-Holland Publishing Company.

McKelvey, V.E. (1974) Potential Mineral Reserves. Resources Policy, December.

Mesarovic, M., and E. Pestel (1974) Multilevel Computer Model of World Development System, Volume II. CP-74-1. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Moiseev, N.N. (1975) Man–Machine Simulation Systems and the Design of Environmental Programs; a Framework for International Research. Lecture to the Faculty of Environmental Studies, York University, Toronto, Ontario, October 1975.

Moiseev, N.N. (1977) The Unity of Natural Scientific Knowledge. Research Memorandum, Economic Research and Systems Planning Group. York University, Toronto, Ontario, September 1977.

Murtagh, B.A., and M.A. Saunders (1978) Large-Scale Linearly Constrained Optimization. Mathematical Programming 14:1.

Naill, R.F. (1972) Managing the Discovery Life-Cycle of a Finite Resource: A Case Study of U.S. Natural Gas. Cambridge, Massachusetts: Sloan School of Management, MIT.

Orchard-Hays, W. (1977) Computerization of the Energy Program's Modelling Effort. IIASA Workshop on Energy Strategies, May 1977. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Propoi, A. (1973) Elementy Teorii Optimalhykh Diskretnykh Protsessov (Elements of the Theory of Optimal Discrete Processes). Moscow: Nauka (in Russian).

Propoi, A. (1976) Problems of Dynamic Linear Programming. RM-76-78. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Propoi, A. (1977) Dual Systems of Dynamic Linear Programming. RR-77-9. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Propoi, A. (1978) Models for Educational and Manpower Planning: a Dynamic Linear Programming Approach. RM-78-20. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Propoi, A. (1979) Dynamic Linear Programming. In L.C.W. Dixon and G.P. Szeyö (Eds.), Numerical Optimization of Dynamic Systems. Amsterdam: North-Holland Publishing Company. (See also IIASA Working Paper WP-79-38, 1979).

Propoi, A., and V. Krivonozhko (1977) The Dynamic Simplex Method. RM-77-24. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Propoi, A., and V. Krivonozhko, (1978) The Simplex Method for Dynamic Linear Programs. RR-78-14. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Propoi, A., and A.B. Yadykin (1974) On a Planning Problem under Conditions of Uncertainty Demand, Parts I and II. Automatika i Telemekhanika, N2 and N3 (in Russian).

Propoi, A.I., and A.B. Yadykin (1975/1976) Parametric Iterative Methods for Dynamic Linear Programming. I, Nondegenerate Case. Automatika i Telemekhanika 12(1975). II, General Case. Automatika i Telemekhanika 1 (1976) (both in Russian).

Propoi, A.I., and A.B. Yadykin (1978) Parametric Quadratic and Linear Programming, Parts I and II. Automatika i Telemekhanika, N2 and N4 (in Russian).

Suzuki, A. (1975) An Extension of the Häfele–Manne Model for Assessing Strategies for a Transition from Fossil Fuel to Nuclear and Solar Alternatives. RR-75-47. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Tomlin, J. (Ed.) (1976) Notes on a Workshop on Energy System Modelling. Report SOL 75-6. Stanford, California: Stanford University Systems Optimization Laboratory.

Zimin, I. (1976a) Mathematical modelling in normative development planning. In E. Billeter, M. Cuenod, and S. Klaczo (Eds.) Overlapping Tendencies in Operations Research, Systems Theory and Cybernetics. Basel and Stuttgart: Birkhäuser Verlag.

Zimin, I. (1976b) Optimal Control Theory Problems with Network Constraints and their Application. RM-76-13. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Zimin, I. (1977) Aggregated Dynamic Model of a National Economy. Discussion Paper, IIASA Workshop on Energy Strategies, May 1977. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Zimin, I. (1980) INTERLINK: An Interindustrial Model for Simulating a Balanced Regional Economic Development and its Links to IIASA's Set of Energy Models. WP-80-95. Laxenburg, Austria: International Institute for Applied Systems Analysis.

ABSTRACTS OF OTHER IIASA PUBLICATIONS

Fedra, K., Mathematical Modelling – a Management Tool for Aquatic Ecosystems? IIASA Research Report RR-81-2, March 1981.
Reprinted from *Helgoländer Meeresuntersuchungen*, Vol. 34, 1980, pp. 221–235.

Mathematical modelling may serve as a rational and powerful tool in the management of complex ecosystems. However, ecosystem models are drastic simplications of the real world. As a rule they are based on a rather incomplete and scattered knowledge of the system in question. Furthermore, ecological systems and in particular marine systems are characterised by a high degree of complexity, spatial and functional heterogeneity, nonlinearity, complex behavioural features such as adaptation and self-organisation, and a considerable stochastic element. Nevertheless, if management is to be based on predictions from mathematical models – and it has to be based on some kind of "model" in at least a broad sense – we need an estimate of prediction accuracy in terms of the management variables and constraints. One possible approach to model uncertainty is a probabilistic interpretation of model predictions, generated by use of Monte-Carlo techniques. Fuzzy data sets and ranges are used. The resulting model response allows the derivation of measures for model credibility. Probability distributions can be computed for certain system states under (un)certain input conditions, representing the effects of insufficient data and structural uncertainty on model-based predictions. Such analysis indicates that prediction uncertainty increases, not only with the uncertainty in the data, but also with increasing "distance" from the empirical conditions, and with time. Present ecosystem models can be a tool for qualitative discrimination between different management alternatives, rather than a credible means for detailed quantitative predictions of system response to a wide range of input conditions.

Clark, W.C., Witches, Floods, and Wonder Drugs: Historical Perspectives on Risk Management. IIASA Research Report RR-81-3, March 1981.
Reprinted from Richard C. Schwing and Walter A. Albers, Jr., editors, *Societal Risk Assessment: How Safe is Safe Enough?* New York: Plenum Press, 1980, pp. 287–314.

Risk is a people problem, and people have been contending with it for a very long time indeed. I extract some lessons from this historical record and explore their implications for current and future practice of risk management.

Socially relevant risk is not uncertainty of outcome, or violence of event, or toxicity of substance, or anything of the sort. Rather, it is a perceived inability to cope satisfactorily with the world around us. Improving our ability to cope is essentially a management problem: a problem of identifying and carrying out the actions that will change the rules of the game so that the game becomes more to our liking.

To cope better is to understand better the nature of risks and how they develop. It is naive and destructive to pretend that such understanding can carry with it the certainties and completeness of traditional science. Risk management lies in the realm of trans-science, of ill-structured problems, of messes. In analyzing risk messes, the central need is to evaluate, order, and structure inevitably incomplete and conflicting knowledge so that the management acts can be chosen with the best possible understanding of current knowledge, its limitations, and its implications. This requires an undertaking in policy analysis, rather than science.

One product of such analyses is a better conceptualization of "feasibility" in risk management. Past and present efforts have too often and too uncritically equated the feasible with the desirable. Results have been both frustrating and wasteful.

Another is an emphasis on the design of resilient or "soft-fail" coping strategies. The essential issue is not optimality or efficiency, but robustness to the unknowns on which actual coping performance is contingent.

The most important lesson of both experience and analysis is that societies' abilities to cope with the unknown depend on the flexibility of their institutions and individuals, and on their capability to experiment freely with alternative forms of adaptation to the risks that threaten them.

Neither the witch hunting hysterics nor the mindlessly rigid regulations characterizing so much of our present chapter in the history of risk management say much for our ability to learn from the past.

Beck, M.B., Hard or Soft Environmental Systems? IIASA Research Report RR-81-4, March 1981.
Reprinted from *Ecological Modelling*, Vol. 11, 1981, pp. 233–252.

Recent trends in lake and stream water quality modeling indicate a conflict between the search for improved accuracy through increasing model size and complexity, and the search for applicability through simplification of already existing models. Much of this conflict turns on the fact that that which can be simulated in principle is simply not matched by that which can be observed and verified in practice. This paper is concerned with that conflict. Its aim is to introduce and clarify some of the arguments surrounding two issues of key importance in resolving the conflict: uncertainty in the mathematical relationships hypothesized for a particular model (calibration and model structure identification); and uncertainty associated with the predictions obtained from the model (prediction error analysis). These are issues concerning the reliability of models and model-based forecasts. The paper argues, in particular, that there is an intimate relationship between prediction and model calibration. This relationship is especially important in accounting for uncertainty in the development and use of models. Using this argument it is possible to state a dilemma, which captures some limiting features of both large and small models.

Häfele, W., A Global and Long-Range Picture of Energy Developments. IIASA Research Report RR-81-8, May 1981.

Reprinted from P.H. Abelson and R. Kulstad, editors, *The Science Centennial Review*, Washington, DC: American Association for the Advancement of Science, 1980, pp. 156–164. The article was originally published in this form in the Centennial Issue of *Science*, Vol. 209, 1980, pp. 174–182.

Most studies of energy supply and demand ignore either global interdependence or the long time spans necessary to adjust to new energy sources. The International Institute for Applied Systems Analysis has therefore studied on a global scale, for seven major world regions, the balance between energy supply and demand for the next 50 years. Reported here are the results for two benchmark scenarios. In the "low" scenario world energy consumption increases from today's 8.2 terawatt-year per year to 22 terawatt-year per year in 2030; in the "high" scenario, consumption increases to 35 terawatt-year per year. The study showed that time will be the limiting constraint in adapting the energy supply infrastructure to changing resource availability; resources will be available until the second half of the next century, but a strong shift will be required to low-grade fossil fuels such as shale oil and tar sands. Each scenario studied indicated increased environmental problems associated with increased use of fossil fuels, and potential geopolitical problems associated with the world distribution of resources.

Miser, H.J., Operations Research and Systems Analysis. IIASA Research Report RR-81-9, May 1981.
Reprinted from P.H. Abelson and R. Kulstad, editors, *The Science Centennial Review*. Washington, DC: American Association for the Advancement of Science, 1980, pp. 121–128. The article was originally published in this form in the Centennial Issue of *Science*, Vol. 209, 1980, pp. 139–146.

The science of man–machine operating systems, which includes operations research and systems analysis, has achieved a substantial body of theory and application over the last 40 years. Its current strength prompts it to attack difficult large-scale problems while challenging the other relevant sciences to unite, not only with each other and operations and systems research, but also with society, to deal with some of the most widespread and important problems of our time.

Weingart, J.W., The Helios Strategy: An Heretical View of the Potential Role of Solar Energy in the Future of a Small Planet. IIASA Research Report RR-81-10, May 1981.
Reprinted from *Technological Forecasting and Social Change*, Vol. 12, 1978, pp. 273–316.

Over the next hundred years there must be a worldwide transition from reliance on fossil fuels to the use of some combination of long-term and abundant primary sources for the production of heat, electricity, and synthetic fuels. The rate at which such options can be developed and employed, as well as the maximum rate at which they can provide energy at a sustained rate, will place important constraints on the rate and limits to growth of other human activities. It is generally argued that only the fission option, in the form of the fast-breeder and high-temperature reactors, can provide the energy required for a livable world, particularly if this means a world of 10 billion people living at the present energy level of Western Europe. However, a careful examination indicates that the use of solar energy, through a menu of technological options, can provide the needs of a world at this scale of energy use, and that this can be accomplished within the constraints of land availability and requirements for energy, materials, and labor. No scientific breakthroughs are required, although a number of these would be helpful, but very substantial engineering advances *are* required, and the transition to such a world-wide system would take no less than a century. However, the feasibility of such large-scale use of solar energy will substantially alter those aspects of the "limits to growth" discussions in which future growth strategies are constrained by available and acceptable energy alternatives. This paper outlines a global solar-energy system considered feasible for more than 10 billion people living at 5 kW per capita.

Nurminski, E.A., Π-Approximation and Decomposition of Large-scale Problems. IIASA Research Report RR-81-11, June 1981.
Reprinted from A. Auslender, W. Oettli, and J. Stoer, editors, *Optimization and Optimal Control*. Berlin: Springer Verlag, 1981, pp. 79–88.

Partial or complete dualization of extremum problems often allows the decomposition of initially large-scale problems into smaller ones with some coordinating program of a moderate size. This idea underlies many known schemes of decomposition and the common difficulty often encountered is the problem of restoring the solution of the primal problem. The main idea of this paper is to present an algorithm for providing an easy way of obtaining the solution of the initial primal problem keeping all advantages of the dual one.

The algorithm described here is based on the particular approximation of the aggregated function representing the decomposed way of solving the extremum problem. This approximation looks like a dual problem and its remarkably simple structure makes it possible to solve a corresponding extremem problem in a few iterations.

Nurminski, E.A., An Application of Nondifferentiable Optimization in Optimal Control. IIASA Research Report RR-81-12, June 1981.
Reprinted from L.C.W. Dixon and G.P. Szegö, editors, *Numerical Optimisation of Dynamic Systems*. Amsterdam: North-Holland Publishing Company, 1980, pp. 137–158.

The problem of optimal control for the nonlinear dynamic system with discrete time is considered. Using a nondifferentiable penalty function it is possible to transform the initial problem into an unconditional one. The special structure of this problem makes it possible to develop the specific method, which is some composition of the gradient-like method of nondifferentiable optimization and the method of coordinate minimization.

BIOGRAPHIES

Lars Bergman, Sweden

Lars Bergman, Assistant Professor at the Stockholm School of Economics, joined IIASA's System and Decision Sciences Area in August 1978. His work involves developing numerically formulated input–output-based or similar models for long-term forecasting and economic policy analysis.

From 1970 to 1975 Professor Bergman was a research assistant and later a research scholar at the Economic Research Institute of the Stockholm School of Economics. From 1975 to 1976 he was a researcher at the Secretariat for Future Studies, Ministry of Education, and in 1976 he returned to the Stockholm School of Economics, receiving his Ph.D. in Economics in 1977.

Professor Bergman's scientific interests include macroeconomic planning, long-term economic policy, in particular energy policy, and economic growth. He has published a monograph on energy consumption and economic growth in Sweden and has also written several articles and reports (in Swedish) to government commissions.

Manfred Breitenecker, Austria

Manfred Breitenecker received his Ph.D. in physics and mathematics in 1960 and his habilitation as docent in theoretical physics in 1975 from the University of Vienna. An assistant at the Institute for Theoretical Physics at the University of Vienna, he has been with IIASA since 1976 as a consultant with the Energy Systems Program. His special interests include mathematical physics, dynamical systems, functional analysis, and teaching.

Hans-Richard Grümm, Austria

Hans-Richard Grümm received his M.S. in mathematical physics in 1970 from the University of Washington, his Ph.D. in the same subject in 1971 from the University of Vienna, and his J.D. in 1976 from the University of Vienna. He is an assistant at the Institute for Theoretical Physics at the University of Vienna. In 1975 he joined IIASA, where his research focuses on the qualitative theory of differentiable dynamical systems as applied to the concept of resilience.

Heinz-Dieter Haustein, GDR

Heinz-Dieter Haustein received his diploma in Industrial Economics in 1955 from the Leningrad Institute of Finance and Economics and his Ph.D. in 1960 and his habilitation in 1964 from the University of Economics, Berlin (East). Deputy Director of the Institute for Management and Head of the Department for Planning and Forecasting of Science and Technology at the Bruno Leuschner University of Economics in Berlin (East), Professor Haustein has been working on the Innovation Management Task of the Management and Technology Area at IIASA since 1979. His research interests include the economic analysis of technology and innovation.

Harry Maier, GDR

Harry Maier graduated in 1957 from the Humboldt University in Berlin (East) and received his Ph.D. in 1961 and his habilitation in 1967 from the same university. He is Deputy Director of the Central Institute of Economic Sciences of the Academy of Sciences of the German Democratic Republic. Professor Maier worked in IIASA's Management and Technology Area from 1978 until 1980. His scientific interests include the economics of education and governmental policy with respect to technology.

N.S.S. Narayana, India

N.S.S. Narayana received his Ph.D. in economics from the Indian Statistical Institute. He was a consultant to the Planning Commission, Government of India, before coming in August 1977 to IIASA, where he worked on the Indian Agricultural Policy and Planning Model of the Food and Agriculture Program. In 1979 Dr. Narayana joined Madras University, India. He rejoined IIASA in April 1981.

Lennart Ohlsson, Sweden

Lennart Ohlsson received his Ph.D. in economics from the University of Uppsala, Sweden, and since 1974 has been a senior research economist at the Expert Group on Regional Studies, Ministry of Industry, Stockholm. In august 1978, Dr. Ohlsson joined the Human Settlements and Services Area at IIASA for one year, during which time he participated in research on changing regional specialization and its employment consequences. He is currently teaching at the Stockholm School of Economics.

Kirit S. Parikh, India

Kirit S. Parikh, Leader of IIASA's Food and Agriculture Program, studied at Gujarat University, the Indian Institute of Technology in Kharagpur, and the Massachusetts Institute of Technology. Formerly Director of the Programme Analysis Group of the Department of Atomic Energy, Bombay, and Professor of Economics at the Planning Unit of the Indian Statistical Institute, New Delhi, Professor Parikh has been author or coauthor of several books, manuals, and articles.

Anatoli Propoi, USSR

Anatoli Propoi was a member of IIASA's System and Decision Sciences Area from July 1976 to August 1979. His work at IIASA concentrated on methods of dynamic optimization, in particular dynamic linear programming, and he cooperated extensively with the Energy Systems and Food and Agriculture Programs and with the Regional Development Task. Dr. Propoi received his diploma in physics (1962) from the Moscow Physico-Technical Institute, and his degrees as Candidate of Sciences (1965) and Doctor of Sciences (1974) from the Institute for Control Sciences. From 1962 to 1976 he was attached to the Institute for Control Sciences, first as a research scholar and later as Head of the Systems Dynamics Laboratory. Since 1976 he has been Head of the Optimization Methods Laboratory of the All-Union Institute for Systems Studies of the State Committee for Science and Technology and the USSR Academy of Sciences. Dr. Propoi's research interests include mathematical programming, optimal control theory, and the optimization of large-scale systems.

Luitpold Ulhmann, FRG

Luitpold Uhlmann graduated in 1961 from the University Erlangen-Nürnberg and later received his Ph.D. from the same university. He worked for IBM from 1962 to 1965; since 1965 he has been a Research Scientist in the IFO–Institute for Economic Research in Munich. His scientific interests include the economic and social problems of technological change.

Igor Zimin, USSR

Igor Zimin joined IIASA in 1974, first with the Methodology Group and later with the Energy Systems Program where he worked on optimization techniques and simulation modeling and their application to economics, energy, and resources. Dr. Zimin graduated from the Moscow Physico-Technical Institute in 1970 and received his degree as Doctor of Physics and Mathematics from the same institute in 1974. From 1970 to 1979 he worked in the area of mathematical economics at the Computer Center of the USSR Academy of Sciences; in 1979 he joined the All-Union Institute for Systems Studies in Moscow. His scientific interests include numerical methods of control theory and mathematical programming, global and regional development modeling, mathematical economics (long-range investment planning and forecasting, resource allocation, and environmental impact of production systems), and applied cybernetics.

IIASA
REPORTS

Volume 3 Number 2 April–June 1981
Completing Volume 3

Published quarterly

CONTENTS

IN FORTHCOMING ISSUES:

Special issue devoted to Advances in Multiregional Demography

Printed by NOVOGRAPHIC, Vienna, Austria.