OPTIMAL GROWTH PATHS WITH EXHAUSTIBLE RESOURCES: AN INFORMATION-BASED MODEL

Robert U. Ayres
International Institute for Applied Systems Analysis
Laxenburg, Austria

RR-87-11
July 1987

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
Laxenburg, Austria

International Standard Book Number 3-7045-0084-4

Research Reports, which record research conducted at IIASA, are independently reviewed before publication. However, the views and opinions they express are not necessarily those of the Institute or the National Member Organizations that support it.

Cover design by Martin Schobel

Printed by Novographic, Vienna, Austria

Summary

An information-based model of the "optimal control" type is developed using concepts from information theory to explore the dynamics of fossil resource exhaustion and the phenomenon of substitution by other forms of capital and technological knowledge. All exhaustible resource stocks and forms of capital (and knowledge) are taken to be equivalent to forms of *information* in the physical sense.

With this background, economic outputs to inputs (productivity) can be defined in common units (bits); and the ratio between them is a natural dimensionless measure of productivity and of technical efficiency, which is a function of the accumulation of knowledge.

The formal model assumes foui stocks: an exhaustible resource stock S, an "ordinary" productive capital stock K_1, and infrastructure capital K_2 (required to utilize renewable resources or some alternate, less available stock of exhaustible resources), and a knowledge stock T. The model permits investments to build up either type of capital or knowledge, simultaneously or independently. The optimal path (which maximizes a discounted utility of long-term consumption) is to invest in whichever type of capital, or knowledge, has the lowest product of shadow price and marginal productivity at any given time.

It is shown that, with optimal policies, the planning period, or cycle, has several distinct phases, with different investment patterns. During the first phase, investment is limited to building up ordinary capital K_1 and knowledge T, and growth of productivity is most rapid; during the second phase, investment shifts to K_2; during the third phase, investment in K_2 continues, along with reinvestment in K_1 to compensate for depreciation; during the fourth phase, there is simultaneous investment in K_1, K_2, and T.

The model has two important qualitative implications: (1) economic growth rates are inherently discontinuous, and (2) the model predicts an evolutionary structural change – viz, the creation of a new sector in response to the progressive exhaustion or obsolescence of previously essential resource or capital stocks. A multiperiod extension is suggested, leading to a tentative explanation for the Kondratieff long-wave phenomenon.

Foreword

This theoretical paper is a contribution to the basic theory of economic growth. It provides for an explicit role for technological change, both independently and in response to the exhaustion of stocks of nonrenewable resources (or, perhaps, obsolescent forms of capital). The paper suggests some interesting explanatory possibilities with regard to "long waves", a theme IIASA has explored for a number of years. It fits well into the TES program, though much of the work was done before the author arrived at IIASA.

T.H. LEE
Program Leader
Technology–Economy–Society

Acknowledgments

Some of the introductory material in this paper is substantially similar to that in an earler, coauthored paper [Ayres and Miller (1980)]. The basic model described herein was also, in part, anticipated by that earlier paper. Unfortunately, the original model formulation was flawed, preventing a straightforward solution or interpretation.

I am very grateful to Evart von Imhoff, Karl-Göran Mäler, and Erno Zalai for helping me to find and eliminate some mathematical bugs. I am also grateful to Thomas Crocker, Ronald Cummings, Allen Kneese, Wilhelm Krelle, Adam Rose, Pieter de Wolff, and an anonymous reviewer for carefully reading the manuscript in an earlier draft and drawing my attention to various deficiences. My thanks should not be construed as an implication that any of the abovementioned individuals is in full agreement with the somewhat controversial information-theoretic approach taken in this paper. Any remaining errors or defects are entirely mine.

Contents

OPTIMAL GROWTH PATHS WITH EXHAUSTIBLE RESOURCES: AN INFORMATION-BASED MODEL

Robert U. Ayres

Introduction

The circumstances of accelerated use, and possible near- or medium-term exhaustion, of fossil energy resources - together with major uncertainties as to the feasibility, cost, and timing of downstream substitutes - constitute a challenge for economic analysis. Several frameworks are possible. The case where technology offers no substitution possibility was examined many years ago by Gray (1914) and Hotelling (1931). Decades later, Nordhaus (1973) considered a variant case in which the supply curve becomes infinite at some finite price, where the so-called "backstop" technology takes over and provides unlimited energy availability. Stiglitz (1974, 1979) assumed technological progress occurs at a constant rate, regardless of policy, indefinitely.

Dasgupta and Heal (1974, 1979) introduced a different twist. In their earlier models, the new technology eliminates the need for the resource, but it arrives exogenously and costlessly at some uncertain time in the future. In more recent work these authors, as well as Kamien and Schwartz (1978, 1982) and others have examined variations in which the new development itself becomes endogenous and costly. In the context of energy analysis, these models largely retain the backstop concept, the focus being on a single, millennial, breakthrough technology and on optimal policy during the interim period. A useful recent summary of the status of this literature has been given by Huettner (1981).

The simple framework adopted in the present paper differs from some of those cited above in several ways. First, no technological millennium, in the Nordhaus sense, is envisaged. Rather, technical progress is identified with a continuously increasing function of technological knowledge, T, which is taken to be an explicit, endogenously determined factor of production. Second, all factors of production, including technological knowledge, are assumed to be forms of condensed or "embodied" information. Information is used in the technical sense introduced by Hartley (1928) and elaborated by Shannon (1948), Brillouin (1953) and others. The substitutability, or interconvertibility, of factors of production follows naturally. Third, following from the above, both resource inputs and outputs of the production process can also be thought of as forms of condensed information and measured in "bits". The ratio of aggregate outputs (e.g., GNP) to inputs thus becomes a natural generalized measure of the state of technology at a given time. (Some features of the present framework were first suggested in Ayres, 1978).

Much more can be said on the last point. It is one to which many economists raise objections, almost reflexively . However, many of the common objections are rooted in intuitive and rather imprecise uses of the concept of information. The following section is intended to provide some explanatory background material on this topic. It can be skipped by any reader who is either (a) already moderately comfortable with standard concepts of information theory, as used by engineers and physicists, or (b) willing to suspend disbelief and accept, for purposes of argument, that all economic quantities (labor, capital, resources, outputs) can be quantitatively measured in the same physical unit ("bits").

One caveat is essential at this point. The assertion that all factors of production, as well as outputs, can be measured in bits does not preclude their also being measurable in value units (e.g., dollars). The two kinds of units need not be proportional, any more than the relative prices of two materials necessarily coincide with their relative masses. The model introduced later does not seek to maximize the absolute information content of economic output. It does seek to maximize the *utility* of that information output. Thus, a subtle and possibly controversial feature of this model is that it assumes the existence of such a utility function, i.e., a consistent relationship between the information embodied in final products and services produced by the economy and the utility thereof. If there is to be a debate, it should probably focus on whether such a utility function can consistently be determined.

Information

Technically speaking, information is a measure of *uncertainty* (Shannon, 1951), of negative entropy or *negentropy* (Brillouin, 1953), or of *distinguishability* or generalized distance (Tribus and McIrvine, 1971). The more distinguishable or nonrandom a subsystem is, the more information it embodies. This is true of telegraphic or telephonic messages, wireless transmissions, photographs, atomic or molecular assemblages, materials, shapes, and physical structures. It is also true of organizations and social systems.

Methods for numerical computation of information content are available for communications applications and for homogeneous physical–chemical systems. Computational schemes can be developed, in principle, for the more complex cases. In general, the information content of a manufactured thing corresponds roughly to the number of symbols or words that would be required to describe it efficiently (e.g., in a computer program).

Solar radiation is information-rich because it is highly distinguishable (in terms of equivalent black-body temperature) from low-temperature background radiation. High-quality metal ores contain information because their composition is highly distinguishable from the surrounding earth's crust; purified metals contain even more information for the same reason. And so on.

Knowledge is a *useful subset* of information that can be regarded as a factor of production. Not all information is knowledge, but all knowledge is information. "Useful", in this context, merely means that it contributes to the production of useful goods and services. A more extended discussion of the relationship between information and knowledge has been included as *Appendix A*.

While knowledge can be assumed to increase, in principle, without physical limit (if one continues investing in R&D), its impact on productivity is assumed to be subject to diminishing returns. Both the assumption of concavity – or diminishing returns – and the assumption that technological knowledge is endogenous to the productive system, are in contrast to views in some of the extant economic growth literature. However, one important notion underlying the approach described in this paper is that natural resources, labor, physical capital, and knowledge are all condensed forms of information and therefore mutually substitutable, within limits to be discussed later.

In fact, it requires no great leap of the imagination, at this point, to interpret physical capital stock as knowledge (i.e., useful information) embodied in material form. Similarly, various skill levels of labor can readily be interpreted as knowledge embodied in human workers. When capital equipment depreciates due to wear and tear, the (useful) information content embodied in its design (form and function) is gradually lost. As a cutting tool loses its physical edge, its distinguishability is obviously decreased, as is its economic productivity.

The interpretation of capital and labor as embodiments of knowledge does not alter the desirability of taking into account the fact that the economic system also depends on a continuing flow of available energy or essergy. Available energy (essergy) is the ultimate resource, in the same sense that all other material resources can be extracted from the earth's crust, in principle, if enough energy is available. Energy (essergy) from the sun is, of course, the ultimate source of all localized negentropic (information) accumulation on the earth. This being so, the solar energy flux is, in effect, a *flux* of information. Similarly, the earth's store of fossil fuels can be regarded as a stock of information. Some of the latter can be captured and embodied by biological and/or technological processes in other, even more condensed forms, such as capital goods or products.

Technical Efficiency and Technological Knowledge

The essential equivalence of resources and energy is widely accepted (e.g., "energy is the ultimate resource"), and the equivalence of useful or available energy and information (negentropy) has already been discussed. Thus, in the final analysis, both economic inputs (resources) and economic outputs (goods and services) can be viewed as forms of information. These forms differ primarily in terms of the extent to which information is embodied in composition, structure, shape or form, and knowledge content or "quality".

The model discussed hereafter assumes that the modern economic system as a whole is a kind of information processor, which continuously converts massive amounts of crude information (negentropy) into a much smaller quantity of refined information. The latter takes the form of knowledge stocks and human services. (There is an obvious analogy here between crude information and crude oil: refined petroleum products have less energy content, but much greater utility, than crude oil.) Both kinds of information flow are measurable in bits/sec. ***The processing efficiency of the economic system can defined as the ratio of information output flux to information input flux.*** This statement is both trivial and truly profound, as will be seen.

It is convenient at this point to introduce a variable $E(T)$, where T is a measure of technological knowledge T, such that E is constrained to the range zero to unity. For reasons that will be clearer subsequently, it is convenient to think of E as a generalized efficiency measure. It is convenient to let

$$E = [1 + \exp(T_0 - T)]^{-1} \tag{1}$$

where T_0 is a large number (by assumption) such that $E = 0.5$ when $T = T_0$. Evidently if T_0 is large, E is very small for small values of T ($T << T_0$) and asymptotically approaches unity for very large T ($T >> T_0$). Solving for T,

$$T = T_0 + \ln(E/1-E) \tag{2}$$

The growth of the stock of technological knowledge T can be presumed, for purposes of the model, to follow a simple law, viz,

$$\dot{T} = J \tag{3}$$

where J is the annual creation (or destruction) of new knowledge. J is a function of time, of course. The rate of embodiment (or fixation) of knowledge in capital, labor, products, etc., is presumably proportional to the rate of acquisition of new knowledge owing to R&D over some prior period.

The productivity measure E satisfies a nonlinear differential equation, viz,

$$\dot{E} = E(1 - E)J \tag{4}$$

where J (previously defined) is the aggregate annual rate of addition to the stock of knowledge. It can be seen that E is an elongated, more or less S-shaped curve. It is exponentially rising, at first, but after passing a point of inflection, it enters a concave region of saturation, asymptotically approaching unity. If J is a constant, it may be noted that the solution to (4) is the familiar logistic curve. This qualitative behavior is, incidentally, characteristic of most individual technology measures over time.

As technical efficiency E asymptotically approaches unity (i.e., progress continues for a very long time), the economic system generates the maximum possible output of final services, per capita, from a given resource (crude information) flux. Nothing whatever is implied about the need for physical materials, as such, since materials can always be recycled from the environment if enough energy is available.

An Optimal Economic Growth Model

I now introduce an explicit optimal growth model incorporating many of the concepts outlined in preceding paragraphs. In this model, it is assumed that labor force is an exogenous variable proportional to population and independent of other economic variables. For the sake of concreteness, let

$$L = \mathrm{b}N \tag{5}$$

where N is the total population.

It is conventional in the economic literature to make the usual Malthusian assumption, for convenience, that population N grows exponentially over time, at a constant rate g. This seems simplistic on biological grounds and unnecessary. A more reasonable assumption seems to be that humans can, and eventually will, regulate their population to the level that can be supported by the physical environment. In fact, the rate of world population growth has declined significantly in the last 20 years. A simple differential equation having roughly the desired asymptotic behavior is as follows:

$$\dot{N} = \mathrm{g}N(1 - N/\bar{N}) \tag{6}$$

where $\bar{N}$ is the maximum population theoretically sustainable by conventional agriculture, given existing world soil characteristics, rainfall, insulation, and topographic conditions (Pearl, 1922; Buringh *et al.*, 1975). Obviously, if humans were able to colonize other planets or grow food in orbiting space colonies, terrestrial limitations would not apply. However, one need not be concerned at present with the numerical value of $\bar{N}$. I will focus attention, subsequently, on aggregate production and consumption, with the understanding that per capita measures are derivable from them.

Next, consider the stock of fixed (constant vintage) invested capital K. The usual assumed accumulation law is

$$\dot{K} = I - \mathrm{d}K \tag{7}$$

where I is the current level of investment and d is the rate of physical depreciation, assumed to be constant, for convenience. The non-negativity of investment $I \geq 0$ implies that fixed capital cannot be consumed, although the stock can decline as a result of depreciation. For internal consistency, K measures the quantity of constant-vintage capital referred to a given vintage year (e.g., 1985).

It is, of course, true that successive technological improvements will tend to increase the capabilities of machines and/or structures built at later times. Thus, a given quantity of constant capital will be equivalent in productive capability to a smaller quantity of current capital, at any future time. This performance improvement reflects the continuous embodiment of new technological knowledge in capital. However, in this model technological knowledge is assumed, for convenience, to be entirely disembodied.

For purposes of this model, it is necessary to define two distinct kinds of capital, K_1 and K_2. By assumption, K_1 is used in the production of final goods and services, while K_2 is used in the direct capture of solar energy (renewable resources). This is assumed, for convenience, to be a capital-intensive activity, though it could also be labor-intensive. Thus, we define

$$K = K_1 + K_2 \tag{8}$$

Similarly, capital investment has two components:

$$I = I_1 + I_2 \tag{9}$$

In the model, crude information (essergy) resources are required to drive economic activity. The quantity of essergy R needed is a function of the total output of goods and services by the economy, $\Pi(K,L,E)$ where $E = E(T)$. Given the view that economic output $\Pi(K,L,E)$ can be measured in terms of information (bits), and resource input R is also a measure of information input, it makes sense to define E as the dimensionless ratio of aggregate information outputs to aggregate essergy inputs R (both measured in bits), viz,

$$E = \frac{\Pi}{R} \tag{10}$$

Note that this ratio is necessarily less than unity because energy becomes increasingly unavailable (i.e., entropy increases) at each stage of the production process from materials extraction to final assembly. As entropy increases, stage by stage, the total information (negentropy) contained in product-plus-environment necessarily decreases. Thus, equation (10) has physical content. In fact, the condition $\Pi/R < 1$ is required by the second law of thermodynamics. Evidently, the essergy resource requirement at any time is precisely

$$R = \frac{\Pi}{E} \tag{11}$$

The supply of essergy R at any given time may come from either of two sources: fossil fuels or some renewable source (such as biomass) originating in the solar flux. In reality, fossil fuels are not free by any means, since they must be extracted, processed, and distributed. However, for purposes of the model, it is interesting to assume the existence of an initial stock S_0 of essergy that can be extracted costlessly at any desired rate until it is exhausted. (Calculating the optimal consumption of such a stock has been called the "cake-eating problem" for obvious reasons.) Useful essergy can also be extracted from the sun, but only in proportion to the amount of capital K_2 invested for that purpose. To be consistent with the viewpoint adopted above, it is also convenient to divide aggregate production itself into two components

$$\Pi = \Pi_1 \{K_1, L, E(T)\} + \Pi_2 \{K_2\} \tag{12}$$

where Π_2 is the output of the "renewable essergy" sector. The latter can be conceptualized as a set of unmanned solar satellites and ground stations, embodying capital K_2, although it might equally well be some other kind of infrastructure. Since the solar-powered utility sector consumes no essergy, (10) and (11) can be simplified by substituting Π_1 for Π. The essergy resource supply at any time can be written

$$R = -\dot{S}_1 + C_2 K_2 \tag{13}$$

where $\dot{S}_1$ (a negative number) is the rate of change of the stock S_1 of fossil essergy and C_2 is a parameter. The numerical subscripts are used to facilitate a later generalization regarding several kinds of alternative essergy stocks, S_1, S_2, S_3, and types of infrastructure, K_2, K_3,... Equation (13) can thus be rewritten to eliminate R

$$-\dot{S}_1 = \Pi_1/E - C_2 K_2 \tag{14}$$

The total amount of exhaustible resources extracted over time is limited to the size of the original stockpile,

$$S_1(0) = -\int_0^\infty \dot{S}_1 \mathrm{d}t = \int_0^\infty (\Pi_1/E - C_2 K_2)\mathrm{d}t \tag{15}$$

using (14).

An assumption adopted in some of the recent energy and economics literature is to treat the resource (essergy) flux R as a state variable (analogous to K) and thus as a factor of production; see, for instance, Hudson and Jorgenson (1974), Allen *et al.* (1976), Manne (1977), and Hogan and Manne (1977). This is compatible with the observed fact that the aggregate essergy flux is roughly proportional to the output of goods and services [equation (11)]. For recent empirical evidence in favor of this view, see Cleveland *et al.* (1984).

However, notwithstanding the fact that essergy is essential for production – a point emphasized by Dasgupta and Heal (1974, 1979) – I believe that to include it as a factor of production on a par with capital and labor would involve some undesirable double-counting of factors. Essergy is both an intermediate and a final good. It is embodied to a small extent in materials; but, for the most part, intermediate essergy is used to operate capital equipment. To a large extent energy (essergy) is a complement, not a substitute, for other factors. [See Berndt and Jorgenson (1973), Berndt and Wood (1977), and Griffin and Gregory (1976).] Hence, to increase the essergy supply without changing capital or labor would have little or no impact on total output. I assume, in effect, that essergy availability is not a limiting factor in the medium term, though it might be a constraint in the very short run (less than 10 years) or the very long run (millions of years).

It is intuitively obvious that investment in capital stock of the second type, K_2, is infeasible until a considerable conventional productive capacity exists. Thus, investments in the earliest period must be either in "ordinary capital" K_1 or knowledge T, depending upon which is more productive at the time. It is not quite obvious which of these two comes first – a problem not unlike "the chicken or the egg" conundrum. Quite possibly, the optimal choice is to invest simultaneously, though in varying ratios, as will be seen later.

Formulation as an Optimal Control Problem

It is appropriate now to introduce a utility function $\mathrm{U}(Y)$ in which Y is aggregate consumption and what is consumed is information in some condensed form. This is the point where many economists may choose to differ with the assumptions in this paper. It is not clear, *a priori*, that such a utility function can be consistently defined. I have already commented briefly on the equivalence of goods to embodied information. Goods, in turn, generate services, which contribute to the maintenance, extension, and enjoyment of life. The "purpose" of life itself is arguable, but human life – after early infancy – seems to be intimately concerned with awareness or consciousness. Awareness, in turn, is impossible without sensory stimulus and response. The fact that a TV set or book "delivers" information services to consumers is obvious. It is perhaps slightly less obvious that a house or car also delivers services (via the senses) and these services are also equivalent to information. In any case, I assume that services constitute a form of information flux, in the same sense that knowledge is a form of information stock.[1]

Having said this, one can make the usual assumption that $\mathrm{U}(Y)$ is strictly concave and twice differentiable. [It follows that $\mathrm{U}'(Y)$ is a decreasing function of Y]. To be consistent, I now define current consumption in terms of production and investment:

$$Y = \Pi_1(K_1,L,T) - I_1 - I_2 - J \tag{16}$$

In any realistic case one can assume that $Y \geq 0$, where $I_1 + I_2 + J < \Pi_1$. It remains to ascertain the optimal path for consumption and the three types of investment.

An optimal consumption-investment policy requires that one maximize an integral (representing welfare) over time, subject to a number of constraints. The expression to be maximized is the following:

$$W = \int_0^{\mathbf{z}} \exp(-\delta t)\mathrm{U}(Y)\mathrm{d}t + \mathrm{a}_1 K_1(\mathbf{z}) + \mathrm{a}_2 K_2(\mathbf{z}) \tag{17}$$

$$+ \mathrm{a}_3 T(\mathbf{z}) + \mathrm{a}_4 S(\mathbf{z})$$

where δ is an assumed intertemporal discount rate or interest rate and $\mathbf{z}$, fixed in advance, is the end of the planning period. In this case, $\mathbf{z}$ is taken to be very large, but finite. The constants a_1, a_2, a_3, and a_4 are inserted to guarantee that the terminal conditions for an optimal solution will be satisfied. They are chosen to put a prohibitively high penalty on negative values of the state variables at the terminal point. Apart from this, a_i need not be specified further (see Arrow, 1968).

The integral W in (17) must be maximized subject to a number of formal restrictions, including the first-order constraints on state variables, viz, (1) or (4), (7), (8), (9), (11), (17), and (13); plus the non-negative investment conditions $I_1 \geq 0$, $I_2 \geq 0$, $J \geq 0$; and the non-negative rate of fossil resource extraction ($\dot{S} \leq 0$). The latter can be expressed in integral form, as in equation (15).

It should be noted that the assumed population growth equation (6) is completely independent of the rest of the system and affects the optimal path of consumption only to the extent that the total of available output must be shared among the entire population at any given time. Note also that the current resource (essergy) flux is *not* a state variable inasmuch as it is absolutely dependent on the total output of goods and services, which defines the demand for resource inputs. It can therefore be eliminated from the equations.

As already pointed out, I have assumed four kinds of "stocks": productive capital (K_1), and energy capital (K_2), knowledge (T), and fossil essergy (S). The technical efficiency variable E is defined by (1) in terms of knowledge T, and vice versa (2). The solutions to the optimization problem are derived in *Appendix B*. It is interesting that the equations are separable and the shadow price trajectories can be derived explicitly, in two cases, up to the time t_s where (in the optimal case) the stock S is exhausted.

Implications

The implications of the model can best be seen by examining the behavior of the four shadow price variables P_{K_1}, P_{K_2}, P_T, and P_S over time. The important thing to observe is that both P_{K_1} and P_T are initially declining functions of time, while P_S and P_{K_2} are initially increasing. The shadow prices at the starting time $t = 0$ need not be identical, but it is shown in *Appendix B* that the optimal investment policy is always to invest in that form of capital whose shadow price is highest. As a stock increases, its shadow price comes down, and conversely.

Now it is worthwhile to examine the behavior of the four shadow prices: P_S [equation (41)], P_{K_2}, [equation (42)], P_{K_1} [equation (44)], and P_T [equation (46)]. From the transversality (boundary) conditions (see *Appendix B*), we have $P_{K_1}(\mathbf{z}) = P_{K_2}(\mathbf{z}) = P_T(\mathbf{z}) = 0$. However, we assume $P_S(\mathbf{z})$ is not constrained to vanish. From the structure of (41), it can be seen that P_S is a monotonically increasing function (exp δt) for times $t \leq t_s$. (It can be shown that, at the time t_s, P_S is discontinuous. For $t > t_s$, P_S is a declining function.) On the other hand, (42) is the product of an increasing function times a decreasing function that becomes zero at time $t = \mathbf{z}$. Thus, it rises monotonically to a maximum, then declines smoothly and monotonically toward the end of the planning horizon.

The expression (44) for P_{K_1} is more complicated, and it has a different behavior. The first term is a monotonically *decreasing* function whenever the exponent is negative, which is true *whenever the marginal productivity of capital* K_1 *is large enough for long enough.* During periods of investment in K_1 ($I_1 > 0$), Q_{K_1} must vanish identically and the integral in the second term (curly brackets) of equation (44) is necessarily positive. Thus, during periods of active investment, P_{K_1} is the product of a decreasing exponential function times a term (in brackets) that starts at a constant, rises rapidly at first (because of the integral over P_S, which is always increasing), but approaches a maximum as the argument of the integral approaches zero. In short, P_{K_1} is, roughly, a declining exponential multiplied by an increasing "S-curve". It is complex enough, however, to have "wiggles", corresponding to periods if (or when) the integral over marginal productivity of capital of type K_1 falls below a critical level, such that the exponent shifts from negative to positive.

It can be seen that the structure of (46) is similar to the structure of (44) and the behavior of P_T is qualitatively similar to that of P_{K_1}.

Thus, at the beginning of the planning period, two of the shadow prices (P_S, P_{K_2}) are increasing, and two of them (P_{K_1}, P_T) are decreasing.

It is common sense to assume that at the beginning of the period P_{K_1} and P_T are large and P_S, P_{K_2} are zero or negligible. (If this were not the case, there could never be any investment in ordinary productive capital K_1 and/or knowledge t_1, without which there could be no economic output from which savings can be extracted for any subsequent investment in alternative resources.) Given the assumption that P_{K_1} and P_T are initially large but declining toward zero, while P_S and P_{K_2} are initially small but increasing, an intersection in trajectories is inevitable.

Whenever two shadow price trajectories intersect, the optimal policy is to shift investment from one to the other form of capital until the curves cross again, and so on. In principle, such investment switches may occur arbitrarily often. The welfare loss that would result from a compromise policy of investing simultaneously in two (or three) types of capital is therefore negligible. Hence, we can safely assume, hereafter, that the two declining shadow prices, P_{K_1} and P_T, are identical, at least during the early period of unrestrained growth.

It can be seen, now, that the optimal sequence of events, in general terms, consists of four distinct phases:

Phase I $(0 \leq t \leq t_1)$ is characterized by declining P_{K_1} and P_T and investment, alternately or simultaneously, in two types of productive capital K_1 and T. During this phase either $I_1 > 0$ or $J > 0$, or both. But during phase I, P_{K_2} is increasing monotonically and $I_2 = 0$. Time t_1 is defined by the condition $P_{K_1} = P_T = P_{K_2}$. It can be shown without difficulty that this must occur before the final exhaustion of fossil resources (t_s).

Phase II $(t_1 \leq t \leq t_2)$ is a transitional period, during which investment is exclusively directed at building up the alternative energy capital K_2. Thus $I_2 > 0$ and $I_1 = 0$, $J = 0$. During this phase P_{K_2} continues to increase, but at a decreasing rate, until it reaches a maximum value, before beginning to decrease. Meanwhile, P_{K_1}, and P_T also change slope. The two shadow prices P_{K_1} and P_T do not coincide during this phase because the stock of productive capital K_1 *depreciates*, whereas the stock of technological knowledge T does not. Thus, on physical grounds one would expect P_{K_1} to increase and P_T to remain constant. Time t_2 is determined by the condition $P_{K_2} = P_{K_1}$.

Phase III $(t_2 \leq t \leq t_3)$ is a second transitional period, during which the optimal investment policy is a combination of K_1 and K_2, either simultaneously or in alternation. This continues until both P_{K_1} and P_{K_2} have declined to the point where they again equal P_T. This defines time t_3.

Phase IV $(t_3 \leq t \leq \mathbf{z})$ is the final phase during which $P_{K_1} = P_{K_2} = P_T$ all decline more or less simultaneously (i.e., in concert) to zero $(I_1 > 0$, $I_2 > 0$, $J > 0)$. It is convenient to equate this point with $\mathbf{z}$, the end of the planning period.

The sequence of phases is shown schematically in *Figure 1*. The implications for economic growth are shown in *Figure 2*. It is important to observe that, during

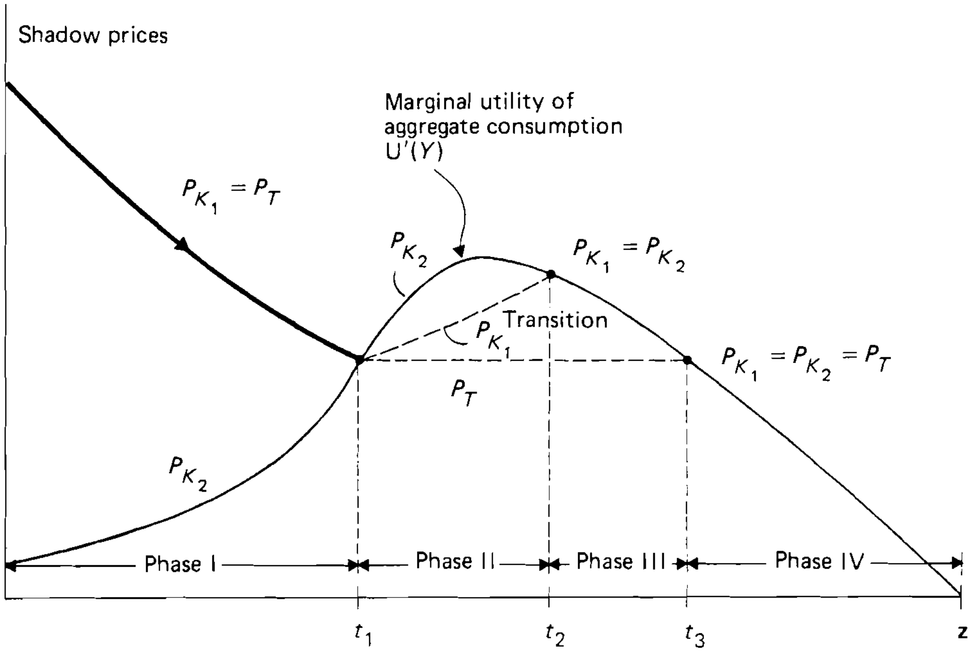

Figure 1. The economic "life cycle".

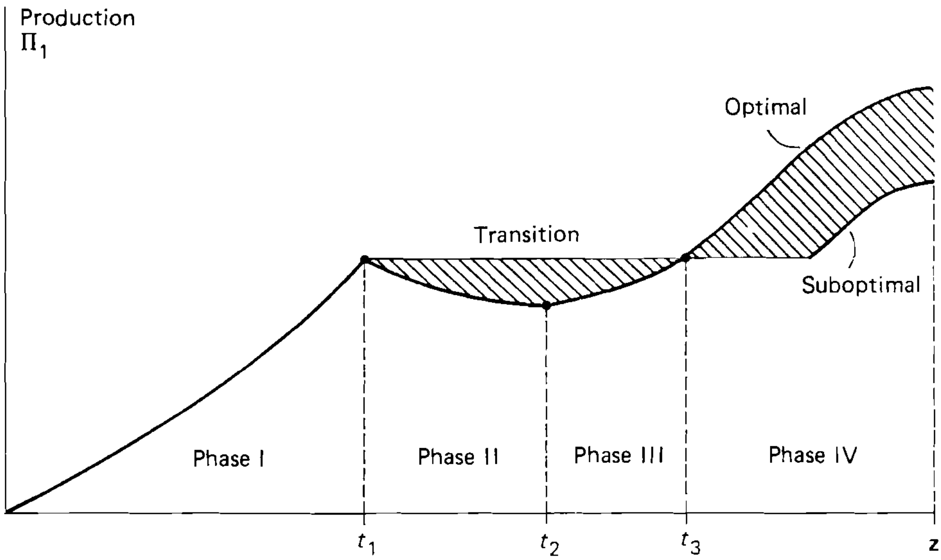

Figure 2. Output of ordinary goods and services (excluding energy infrastructure capital).

Phase II, while investment is devoted exclusively to the build up of K_2 (energy capital), the stock of ordinary capital K_1 is actually declining, whence total output of ordinary goods and services Π_1 must also decline. A feasible (but suboptimal) policy is to invest simultaneously in K_1 and K_2 (as in Phase III), so as to just compensate for depreciation of K_1. It might even be feasible to maintain a slow rate of increase in Π_1 by investing simultaneously in all three forms of capital (as in Phase IV). Obviously, either of these policies would stretch out the transition, resulting in a somewhat lower level of output in Phase IV and a lower final level (at $\mathbf{z}$).

There are two significant implications of this result. First, a long-run optimal policy [given the specification of welfare in (22)] is *inherently discontinuous*, at least as regards economic growth.[2] It is not optimal to invest in energy capital K_2 at early stages of the economic life cycle while the stock of fossil resources S is still large, and it is not optimal to invest in K_1 during the first part of the transition; finally, it is not optimal to invest in knowledge T after time t_1 until K_1 has been restored to its previous level. It follows from the shifts in optimal investment policy that economic growth will tend to follow an irregular path. In particular, sharp discontinuities in growth rate would be experienced, including a change from positive growth rate (slope) to negative growth rate at time t_1.

The reason for the discontinuities on the optimal path has been characterized by Arrow as "myopia". It must be remembered that the control model implicitly postulates an investment decision algorithm based on shadow prices of various types of capital. In principle, these variables are continuously monitored in real time, and investments for the next period are shifted to whatever form of capital currently corresponds to the largest shadow price.

Of course, growth rate discontinuities in the real world tend to be painful (and a more realistic utility function might attach higher utility to paths exhibiting less discontinuous behavior, *ceteris paribus*). An easier way out of the difficulty (suggested by Arrow) is to postulate a "central planner" with some foresight. The planner would be allowed to smooth over potential discontinuities by starting each investment shift somewhat early and extending it beyond the point of theoretical intersection of shadow prices. There would be a small welfare loss relative to the pure (myopic) optimum, but the planner could try to balance the welfare loss with the pain (loss) due to discontinuities. Probably, the planner would use an optimal control model in a simulation mode.

Second, the model inherently accounts for (i.e., predicts) structural changes in the economy. In the simple version described above, *a new sector* is created beginning at time t_1. In the generalized version, discussed later, it can be seen that this sector-creation process can be repeated many times. It may be noted that this seems to be a completely new feature of the present growth model. Earlier equilibrium-type growth models of Harrod (1936), Domar (1956), or von Neumann (1945) are not compatible with structural change of the kind predicted here.

A Multiperiod Generalization

On reflection, the rather specialized model analyzed above can probably be generalized quite easily. The key feature of the model, as described, is the exhaustion of a stock of available "fossil" essergy and the buildup of a specialized stock of capital, K_2, whose only function is to permit the economic system to exploit renewable (solar) energy. However, the optimal path for economic growth would be unchanged if K_2 were interpreted, instead, as a stock of "infrastructure" capital required to enable the use of a *different* (less readily available) stock of exhaustible essergy. For analytic convenience, it was assumed that the building of this capital stock requires "ordinary" capital and labor, but that, once built, each unit of such infrastructure generates a continuous but decreasing flow of essergy throughout its useful life without additional labor. This is a reasonable description of a solar satellite, as noted earlier, or a hydroelectric plant. It is also a fairly realistic characterization of an oil or gas field, after the drilling is completed and the pipelines are in place.

Given this generalized interpretation, Phase IV of the one-period model would effectively become Phase I of a subsequent cycle. At some time perhaps after t_3, but certainly before $\mathbf{z}$, the "planner" would have to assess the magnitude of the second kind of resource stock (call it S_2), which need not be accurately known at the time of the initial plan, and identify the next specialized type of infrastructure capital, K_3, and its annual essergy yield, C_3. A new optimal plan would then be generated for the next period. *Figures 3* and *4* suggest, in schematic terms, how a multiperiod version of the model can be expected to behave.

It is undeniable that *Figure 4* bears some resemblance to the so-called Kondratieff long wave. Many economists still doubt that the cycle is "real". However, if the model described in this paper is at all realistic, a wave-like behavior should exist, though the periodicity need not be constant, and the transitions would be fuzzier (less "bang-bang") than suggested by the simple model.

It is also undeniable that, over the past 200 years of rapid industrialization, there has been a series of fairly dramatic shifts in dominant energy (essergy) technology, from wood (charcoal) to coal, then to petroleum and electricity (derived primarily from fossil fuels), and currently to natural gas and/or nuclear power. The sequence of substitutions is shown schematically in *Figure 5*.

According to the logic of the model, a period of slow growth in ordinary productivity should have occurred during the transition from wood to coal (1780s in the UK, 1880s in the USA); again during the transition from coal dominance to oil dominance (the 1930s?); and finally during the transition from oil to gas (the 1980s?). This is a fascinating speculation, to be sure, but too heavy a burden to lay on such a simple model at this stage. Nevertheless, it is interesting to note that the behavior predicted by the model is, at least, qualitatively, consistent with some aspects of historical experience.

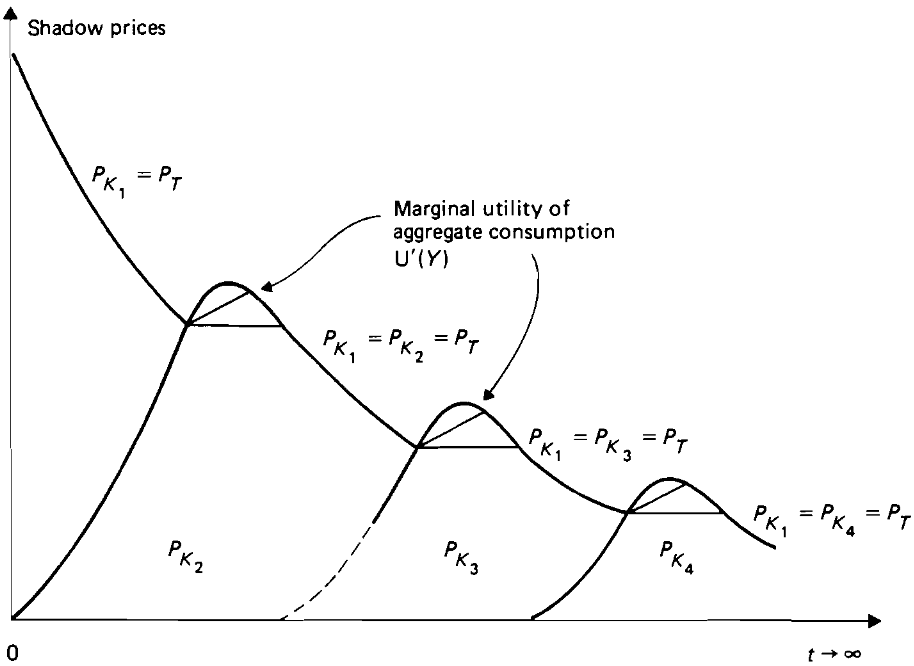

Figure 3. Generalized economic cycle.

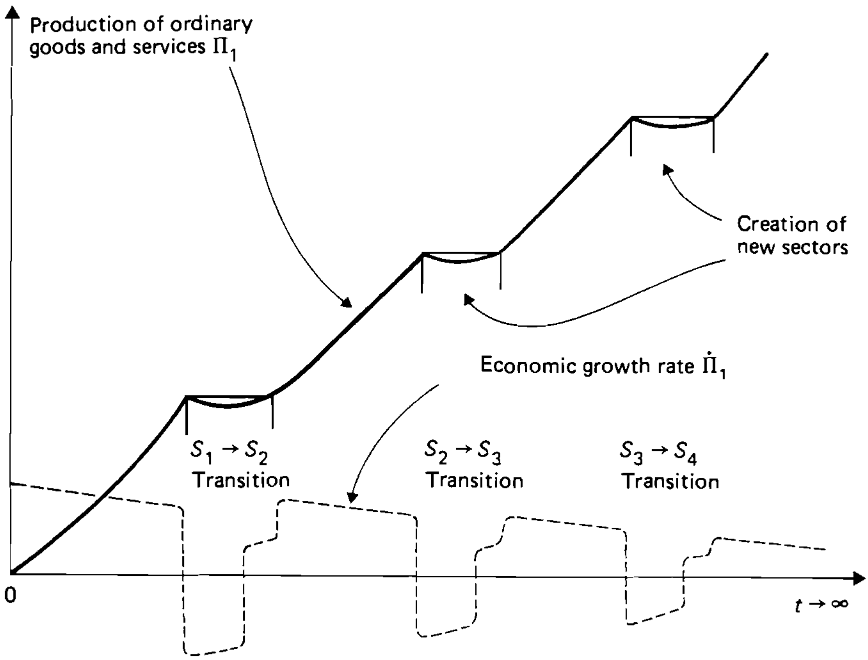

Figure 4. Generalized pattern of growth.

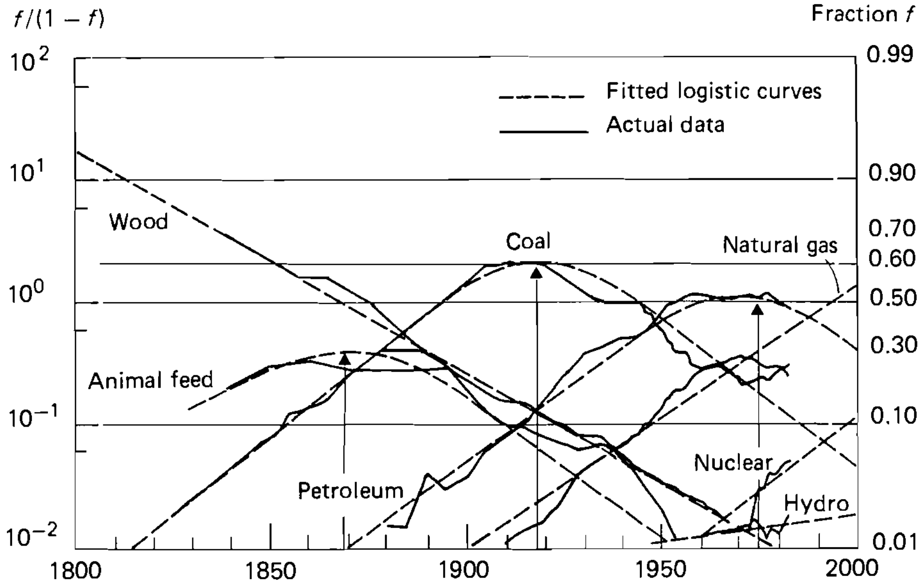

Figure 5(a). USA – primary energy substitution, 1800–2000 (Nakicenovic, 1986).

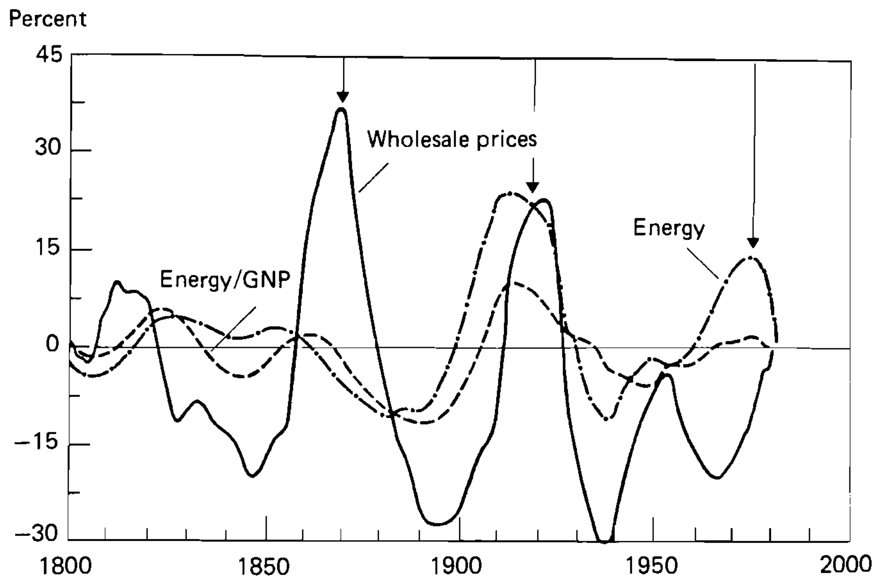

Figure 5(b). USA – energy, energy/GNP, and wholesale prices, 1800–1980 (Nakicenovic, 1986). Note that price peaks correspond to peak shares in *Figure 5(a)*.

Conclusion

The picture is still too crude to adequately reflect what happens in the real world, of course. One obvious oversimplification is the implicit assumption that each essergy source is homogeneous in grade, with constant capital/output ratio (or yield factor $C_2, C_3, \ldots$) over its lifetime. This is unrealistic, of course, and real resources are quite heterogeneous. Moreover, it has been shown by Herfindahl (1967), among others, that it is optimal to utilize the highest grades of ore first. As a consequence, the quality or grade of the remaining stock of any fossil resource tends to decline over time, which implies that more and more economic effort must be devoted to extraction and refining activities over time. This means that the surplus for consumption or reinvestment lags increasingly over time, in comparison with what it would be in the idealized case illustrated by *Figure 4*.

Heterogeneity of actual resource stocks, together with heterogeneity of uses, explains why it can be optimal to exploit more than one different type of resource at the same time, as occurs in the real world.[3] Inhomogeneity and heterogeneity in the system undoubtedly help to smooth out, to some extent, the sharp discontinuities in economic growth rate shown in *Figure 4*. But it is not likely that smoothing due to heterogeneity could totally eliminate the occurrence of changes in growth rate from time to time owing to periodic shifts in investment from one major resource infrastructure to another.

A more penetrating criticism of the present model might be that it is based on the assumption of a single utility function for society as a whole. It is certain that humans and organizations do not in general attempt to specify a utility function for decision-making. In fact, most do not utilize any optimizing methodology, formal or otherwise. Even if firms or individuals can be assumed to behave like rational "utility maximizers", in the sense of von Neumann and Morgenstern (1944), it is unclear that the combined behavior of many independent individual decision-makers would result in overall economic behavior equivalent to that of a single utility-maximizing entity.

Thus, the realism of any such model as this is open to question on several points. Nevertheless, the model seems to capture two important but hitherto elusive aspects of macroeconomic behavior. This would appear to justify further investigations, both at the theoretical and empirical levels.

Appendix A:

Information and Knowledge

In fact, one can clearly identify and distinguish at least three distinct forms of information. There is an obvious analogy with the distinct forms of matter (solid, liquid, and gas), although I do not pursue it further here.

(1) *Disembodied* information is associated with the temperature or spectral characteristics of incoherent electromagnetic or thermal radiation (energy). It is quantitatively proportional to the available useful work (or essergy) content of the energy flux.
(2) Information is *embodied* in the (average) state and chemical composition of unstructured matter, whether gaseous, liquid, or solid, or in the physical microstructure of a crystal or glass.
(3) Information is embodied in the form or shape of a solid medium (two or three dimensions) or in the structure of a macromolecule (such as DNA).

The first two categories are essentially thermodynamic. Explicit rules for computing each type of information content in quantitative terms have been formulated. Note that the third category includes information as we normally use the term, e.g., a photograph, symbols on a printed sheet of paper, a magnetized tape, a precision gear, or a pattern of impurities in a silicon chip. Information of the third kind can be, and often is, transmitted via telecommunications channels, converted from one form (e.g., analog data) to another equivalent form (e.g., digital data) and "processed" by computers.

Note that the third kind of information can only be stored and processed (i.e., utilized) by living organisms and/or material devices that also embody information of the second kind. Moreover, all such organisms and devices require a flux of available useful work (essergy) for their maintenance. Thus, information of the third kind is, in some sense, the essence or condensate of a much larger quantity of information of the first and second kinds.

Knowledge can perhaps be thought of as a fourth kind of information or as the "useful" component of information of the third kind. It has been suggested that knowledge is the minimum information required to decode a message or to reproduce forms or patterns. If this is true, knowledge is a form of information embodied in a decoder or copying machine, or possibly in a living reproductive cell or a brain. Knowledge is therefore literally undefinable in the absence of a supporting material system. The more knowledge is embodied in the decoder, the less information needs to be transmitted to reproduce the original message, or object.

There is no general means of computing the minimum information requirement to reproduce an object, except for objects themselves defined in terms of computer languages. In this context, it is noteworthy that there is a computer science literature on algorithmic information theory; see, for example, Chaitin (1978). Although quantitative formulae are lacking in general, it is safe to assume that the knowledge component of stored or transmitted information of the third kind is normally quite small, compared to the total amount of information of all kinds that must be mobilized to store or transmit it. In other words, much form and structure information is actually redundant. It follows, incidentally, that while the amount of thermodynamic information (of the second kind) that can be extracted each year from all sources (fossil fuels plus solar flux) is indeed limited, this in itself imposes no practical limitation on the rate of accumulation of human knowledge relevant to the production of goods or services.

Appendix B:

Solution to the Optimization Problem

To solve the optimization problem stated above (following Takayama, 1974), we define a present value Hamiltonian system with three “controls” I_1, I_2, and J:

$$
\begin{aligned}
H = {} & \lambda_0\Big[\mathrm{U}\Big(\Pi_1 - I_1 - I_2 - J\Big) + \mathrm{a}_1\dot{K}_1 + \mathrm{a}_2\dot{K}_2 + \mathrm{a}_3\dot{T} + \mathrm{a}_4\dot{S}_1\Big] \\
& + \hat{P}K_1\Big(I_1 - \mathrm{d}_1K_1\Big) + \hat{P}_{K_2}\Big(I_2 - \mathrm{d}_2K_2\Big) + \hat{P}_TJ \\
& - \hat{P}_S\left(\frac{\Pi_1}{E} - C_2K_2\right) + Q_{K_1}I_1 + Q_KI_2 + Q_TJ + Q_SS_1
\end{aligned}
\tag{18}
$$

It can be shown without difficulty that λ_0 can be set equal to unity without loss of generality. Moreover, the three terminal conditions are automatically satisfied by defining $P_{K_1} = \hat{P}_{K_1} + \mathrm{a}_1$, $P_{K_2} = \hat{P}_{K_2} + \mathrm{a}_2$, $P_T = \hat{P}_T + \mathrm{a}_3$, and $P_S = \hat{P}_S - \mathrm{a}_4$. This yields the simpler equivalent Hamiltonian:

$$
\begin{aligned}
H = {} & \mathrm{U}\Big(\Pi_1 - I_1 - I_2 - J\Big) + P_{K_1}\Big(I_1 - \mathrm{d}_1K_1\Big) \\
& + P_{K_2}\Big(I_2 - \mathrm{d}_2K_2\Big) + P_TJ - P_S\left(\frac{\Pi_1}{E} - C_2K_2\right) \\
& + Q_{K_1}I_1 + Q_{K_2}I_2 + Q_TJ + Q_SS
\end{aligned}
\tag{19}
$$

The co-state variables P_{K_1}, P_{K_2}, P_T, and P_S are canonical conjugates of K_1, K_2, T, and S_1, respectively. They are usually interpreted as shadow prices of the corresponding stocks K_1, K_2, S_1, and T. The Lagrange multipliers Q_K, Q_T, and Q_S are zero or positive, but the products Q_{K_1}, I_1, Q_{K_2}, I_2, Q_TJ, and Q_SS_1 are all identically zero. Thus, introducing the non-negativity constraints:

$$
\begin{aligned}
Q_{K_1} &= 0 \text{ whenever } I_1 > 0\text{; otherwise } Q_{K_2} \geq 0 \\
Q_{K_2} &= 0 \text{ whenever } I_2 > 0\text{; otherwise } Q_{K_2} \geq 0 \\
Q_T &= 0 \text{ whenever } J > 0\text{; otherwise } Q_T \geq 0 \\
Q_S &= 0 \text{ whenever } S > 0\text{; otherwise } Q_S \geq 0
\end{aligned}
$$

Two other non-negativity constraints could be included: $-\dot{S} \geq 0$ (resources are never put back into the ground) and $I_1 + I_2 + J < \Pi_1$ (investment never exceeds current production). However, the constrained and unconstrained solutions are essentially identical.

The first three Euler–Lagrange equations for an optimal path are obtained by partially differentiating the Hamiltonian (19) with respect to I_1, I_2, and J, respectively:

$$\frac{\partial H}{\partial I_1} = 0 = -\mathrm{U}'(Y) + P_{K_1} + Q_{K_1} \tag{20}$$

$$\frac{\partial H}{\partial I_2} = 0 = -\mathrm{U}'(Y) + P_{K_2} + Q_{K_2} \tag{21}$$

$$\frac{\partial H}{\partial J} = 0 = -\mathrm{U}'(Y) + P_T + Q_T \tag{22}$$

It follows from (20), (21), and (22) that

$$P_{K_1} + Q_{K_1} = P_{K_2} + Q_{K_2} = P_T + Q_T = \mathrm{U}'(Y) \tag{23}$$

where $\mathrm{U}'(Y)$ is the marginal utility of aggregate consumption. It follows from (23) and the non-negativity conditions that the optimal investment rule is to invest in that type of capital with the largest shadow price. This can be demonstrated by assuming the contrary. For instance, let $P_{K_1} < P_{K_2}$ but assume simultaneous investment in both, i.e., $I_1 > 0$ and $I_2 > 0$ at the same time. Then from the non-negativity conditions $Q_{K_1} = Q_{K_2} = 0$. From (23) it would follow that $P_{K_1} = \mathrm{U}'(Y)$ and that $P_{K_2} = \mathrm{U}'(Y)$. But this is not possible, by assumption that $P_{K_1} < P_{K_2}$.

Taking this line of reasoning further, one can now derive the following expressions for the Qs:

$$Q_{K_1} = \begin{cases} 0 & \text{if} \quad P_{K_1} > P_{K_2}, P_T \\ P_{K_2} - P_{K_1} & \text{if} \quad P_{K_2} > P_{K_1}, P_T \\ P_T - P_{K_1} & \text{if} \quad P_T > P_{K_1}, P_{K_2} \end{cases} \tag{24}$$

$$Q_{K_2} = \begin{cases} 0 & \text{if} \quad P_{K_2} > P_{K_1}, P_T \\ P_{K_1} - P_{K_2} & \text{if} \quad P_{K_1} > P_{K_2}, P_T \\ P_T - P_{K_2} & \text{if} \quad P_T > P_{K_1}, P_{K_2} \end{cases} \tag{25}$$

$$Q_T = \begin{cases} 0 & \text{if} \quad P_T > P_{K_1}, P_{K_2} \\ P_{K_1} - P_T & \text{if} \quad P_{K_1} > P_{K_2}, P_T \\ P_{K_2} - P_T & \text{if} \quad P_{K_2} > P_{K_1}, P_T \end{cases} \tag{26}$$

The co-state variables P_{K_1}, P_{K_2}, P_T, and P_S, together with the corresponding state variables K_1, K_2, T, and S, satisfy the following canonical system of differential equations, which are conditions for a solution:

$$\frac{\partial H}{\partial P_{K_1}} = \dot{K}_1 \tag{27}$$

$$\frac{\partial H}{\partial P_{K_2}} = \dot{K}_2 \tag{28}$$

$$\frac{\partial H}{\partial P_T} = \dot{T} \tag{29}$$

$$\frac{\partial H}{\partial P_S} = \dot{S} \tag{30}$$

$$\frac{\partial H}{\partial K_1} = -\left(\dot{P}_{K_1} - \delta P_{K_1}\right) \tag{31}$$

$$\frac{\partial H}{\partial K_2} = -\left(\dot{P}_{K_2} - \delta P_{K_2}\right) \tag{32}$$

$$\frac{\partial H}{\partial T} = -\left(\dot{P}_T - \delta P_T\right) \tag{33}$$

$$\frac{\partial H}{\partial S} = -\left(\dot{P}_S - \delta P_S\right) \tag{34}$$

To solve the set of eight differential equations (27–34) we need eight constants of integration. These are determined by so-called transversality conditions. For the four state variables K_1, K_2, T, and S, it is reasonable and sufficient to fix initial values at time $t = 0$. The initial values can be zero or finite. Except for P_S, the corresponding co-state variables must be fixed at the terminal point $t = \mathbf{z}$. Here it is reasonable (though not necessary) to assume

$$P_{K_1}(\mathbf{z}) = P_{K_2}(\mathbf{z}) = P_T(\mathbf{z}) = 0 \tag{35}$$

However, if we specify $S(\mathbf{z}) = 0$, then $P_S(\mathbf{z}) > 0$ and conversely. The case of a declining, but still non-zero, resource stock would imply $P_S(\mathbf{z}) > 0$, with the useful simplification that $Q_S = 0$ for all $t \leq t_s$. The following derivation is based on the simplification that $\mathbf{z} = t_s$, whence $Q_S = 0$ at all times $t \leq \mathbf{z}$. In the more general case where $t_s < \mathbf{z}$, it can be shown that P_S is discontinuous at t_s and declines for $t_s < t < \mathbf{z}$. In this period Q_S is non-zero and must be determined by using the condition $\dot{S} = 0$, which implies (from 14) that $C_2 K_2 = \Pi_1/E$.

It is interesting to note, by the way, that the condition of vanishing shadow prices (35) implies that the marginal utility of consumption $U'(Y)$ also declines to zero (by 23), which means that a consumption plateau is finally reached at $t = \mathbf{z}$. When the indicated differentiations of the Hamiltonian are carried out, the results are set of four first-order differential equations for the shadow prices P_{K_1}, P_{K_2}, P_T, and P_S as follows:

$$0 = \dot{P}_{K_1} - \left[\delta + d_1\right] P_{K_1} + U'(Y)\left[\frac{\partial \Pi_1}{\partial K_1}\right] - P_S\left[\frac{1}{E}\,\frac{\partial \Pi_1}{\partial K_1}\right] \tag{36}$$

$$0 = \dot{P}_{K_2} - \left[\delta + d_2\right] P_{K_2} + C_2 P_S \tag{37}$$

$$0 = \dot{P}_T - \delta P_T + U'(Y)\left[\frac{\partial \Pi_1}{\partial T}\right] - P_S\frac{\partial}{\partial T}\left[\frac{\Pi_1}{E}\right] \tag{38}$$

$$0 = \dot{P}_S - \delta P_S + Q_S \tag{39}$$

It should be pointed out that (38) assumes that output Π_1 is explicitly dependent on T, but that there is no implicit dependence through K_1 or K_2. In other words,

$$\partial K_1/\partial T = \partial K_2/\partial T = 0. \tag{40}$$

This reflects the fact that K_1 and K_2 are pure measures of the quantity of capital. Improvements in the quality of capital and labor are reflected by increases in T alone.

The most general solution of (39) is

$$P_S(t) = e^{\delta t}\left[P_S(0) - \int_{t_s}^{t} Q_S(t')\,dt'\right] = P_S(0)e^{\delta t} \tag{41}$$

since $Q_S = 0$ holds for all $t < t_s$.

The next step is to substitute (41) into (37) and solve. The result is

$$P_{K_2} = \exp(\delta + d_2)t \left[P_K(0) - C_2 \int_0^t P_S(t') \exp(d_2 t')dt' \right] \tag{42}$$

which rises to a maximum value (when $\dot{P}_{K_2} = 0$) and then falls, becoming negative when the term in big square brackets becomes negative. The initial value of $P_K(0)$ must be chosen large enough such that P_{K_2} vanishes at $t = z$, as required by (35).

To solve (36) and (38) we can again substitute (41) and also use (23). Equation (36) becomes

$$O = \dot{P}_{K_1} - \left(\delta + d_1 - \frac{\partial \Pi_1}{\partial K_1}\right) P_{K_1} - \left(\frac{\partial \Pi_1}{\partial K_1}\right)\left(\frac{P_S}{E} - Q_{K_1}\right) \tag{43}$$

with the general solution

$$P_{K_1} = \exp\left[\int_0^t \left(\delta + d_1 - \frac{\partial \Pi_1}{\partial K_1}\right) dt'\right] \left\{P_{K_1}(0) + \int_0^t \left(\frac{P_S}{E} - Q_{K_1}\right) \frac{\partial \Pi_1}{\partial K_1} \exp\left[-\int_0^{t'} \left(\delta + d_1 - \frac{\partial \Pi_1}{\partial K_1}\right) dt''\right] dt'\right\} \tag{44}$$

where Q_{K_1} is given by (24). Similarly (38) becomes

$$0 = \dot{P}_T - \left(\delta - \frac{\partial \Pi_1}{\partial T}\right) P_T - \frac{\partial}{\partial T}\left(\frac{\Pi_1}{E}\right) P_S + \left(\frac{\partial \Pi_1}{\partial T}\right) Q_T \tag{45}$$

and has the solution

$$P_T = \exp\left[\int_0^t \left(\delta - \frac{\partial \Pi_1}{\partial T}\right) dt'\right] \left\{P_T(0) + \int_0^t \left[P_S \frac{\partial}{\partial T}\left(\frac{\Pi_1}{E}\right) - Q_T \frac{\partial \Pi_1}{\partial T}\right] \exp\left[\int_0^{t'} \left(\delta - \frac{\partial \Pi_1}{\partial T}\right) dt''\right] dt'\right\} \tag{46}$$

and Q_T is given by (26). When (24) is inserted in (44) to eliminate Q_{K_1}, the result is an integral equation. The same is true when (26) is substituted into (46). It can be verified without much difficulty that (44) and (46) are well-behaved for reasonable values of the parameters. In particular, the shadow price of productive capital, P_{K_1}, is a (generally) decreasing function of time – as it should be – as long as the marginal productivity of capital $\partial\Pi_1/\partial K_1$ is greater than the sum of pure utility discount rate (if any) plus the depreciation rate, i.e.,

$$\frac{\partial\Pi_1}{\partial K_1} > \delta + \mathrm{d}_1$$

Similarly, the shadow price of technological knowledge P_T is a generally decreasing function of time, provided the marginal productivity of knowledge exceeds the discount rate

$$\frac{\partial\Pi_1}{\partial T} > \delta$$

Note that Q_{K_1} and Q_T are, respectively, nonzero when, and only when, the corresponding investment terms (I_1 and J) vanish. The effect of nonzero values of Q_{K_1} and Q_T is to decrease (or even reverse) the rate of decline of P_{K_1} and P_T, again as one would expect. In fact, Q_{K_1} and Q_T, appearing in equations (44) and (46), act as negative feedback stabilizers, in effect. They vanish at points where the shadow price trajectories intersect and increase as they diverge.

Actual solutions of equations (44) and (46) require forward integration with assumed starting values of P_{K_1}, P_T to $t = \mathbf{z}$, followed by a set of successive corrections until the terminal conditions are satisfied.

Notes

[1] I am indebted to T. Vasko for pointing out that the Soviet scientist Academician V.A. Trapeznikov made essentially the same argument in 1966 (reference not available).

[2] It must be pointed out that the phase transition is *not* discontinuous in the sense that the consumable resource runs out suddenly. Nor is there any discontinuity in resource price at the point where the initial substitution of the alternative of essergy resource (renewable or not) for the depletable resource. This transition begins at a point where the shadow prices are equal. An extended discussion of this issue can be found in Tietenberg (1984: Chapter 6). I am indebted to an anonymous reviewer for calling my attention to this point.

[3] The "grade" of a resource can be defined only in relation to a specific use. Thus, coal is a very low-grade resource in terms of providing liquid fuels for automotive vehicles or aircraft, and it is scarcely better in terms of providing gas for household heating. On the other hand, coking coal is a very high-grade resource for purposes of smelting iron ore.

References

Allen, E.L. *et al.* (1976), *U.S. Energy And Economic Growth, 1975–2010*, Publication ORAU/IEA-76-7 (Institute for Energy Analysis, Oak Ridge Associated Universities, Oak Ridge, TN).

Arrow, K.J. (1968), Applications of control theory to economic growth, in Dantzig, G. and Veinott, A. [eds], *Mathematics of the Decision Sciences* (American Mathematical Society, place?).

Ayres, R.U. (1978), *Resources, Environment and Economics: Applications of the Materials/Energy Balance Principle* (Wiley, New York, NY).

Ayres, R.U. and Miller, S.M. (1980), The role of technological change and resource constraints on an optimal growth path, *Journal of Environmental Economics & Management* (7): 353–371.

Berndt, E.R. and Jorgenson, D.W. (1973), Production structure, in *U.S. Energy Resources and Economic Growth* (Ford Foundation Energy Policy Project, October).

Berndt, E.R. and Wood, D.O. (1977), *Engineering and Econometric Approaches to Industrial Energy Conservation and Capital Formation: A Reconciliation*, Energy Laboratory Working Paper MIT-EL-77-040WP (Massachusetts Institute of Technology, Cambridge, MA).

Brillouin, L. (1953), Negentropy principle of information, *Journal of Applied Physics* **24**(9): 1152–1163.

Buringh, P., Van Heemst, H., and Staring, G. (1975), *Computation of the Absolute Maximum Food Production of the World* (Agricultural University, Wageningen, Netherlands).

Chaitin, G. (1978), Toward a mathematical definition of life, in Levine, R. and Tribus, M. [eds], *The Maximum Entropy Formalism* (MIT Press, Cambridge, MA).

Cleveland, C.J., Costanza, R., Hall, C.A.S., and Kaufmann, R. (1984), Energy and the US economy: A biophysical prespective, *Science* 225 (31 Aug): 890–897.

Dasgupta, P. and Heal, G. (1974), The optimal depletion of exhaustible resources, *Review of Economic Studies*.

Dasgupta, P. and Heal, G. (1979), *Economic Theory and Exhaustible Resources* (Cambridge University Press, Cambridge Economic Handbooks, Cambridge, UK).

Domar, E.D. (1956), Capital expansion, rate of growth and employment, *Econometrica* **14**: 137–147.

Gray, L.C. (1914), Rent under the assumption of exhaustibility, *Quarterly Journal of Economics* **28**: 466–489.

Griffin, J.M. and Gregory, P.R. (1976), An intercountry translog model of energy substitution responses, *American Economic Review* **66**: 845–857.

Harrod, R.F. (1936), *The Trade Cycle* (Oxford University Press, Oxford, UK).

Hartley, R.V.L. (1928), Transmission of information, *Bell System Technical Journal* **7**.

Herfindahl, O. (1967), Depletion and economic theory, in Gaffney, M. [ed], *Extractive Resources And Taxation* (University of Wisconsin Press, Madison, WI).

Hogan, W.W. and Manne, A.S. (1977), Energy–economy interactions: The fable of the elephant and the rabbit, in Hitch, C.J. [ed], *Modeling Energy-Economy Interactions: Five Approaches* (Resources for the Future, Research Paper R5, Washington, DC).

Hotelling, H. (1931), The economics of exhaustible resources, *Journal of Political Economy* **39**: 137–175.

Hudson, E. and Jorgenson, D. (1974), US energy policy and economic growth, 1975–2000, *Bell Journal of Economics & Management Science* **5**: 461–514.

Huettner, D.A. (1981), Energy, entropy and economic analysis: Some new directions, *The Energy Journal* **2**(2): 123–130.

Kamien, M. and Schwartz, N. (1978), Optimal exhaustible resource depletion with endogenous technical change, *Review of Economic Studies* **45**.

Kamien, M. and Schwartz, N. (1982), *Market Structure and Innovation* (Cambridge University Press, Cambridge Surveys of Economic Literature, Cambridge, UK).

Manne, A.S. (1977), ETA-Macro, in Hitch, C.J. [ed], *Modeling Energy-Economy Interactions: Five Approaches* (Resources for the Future, Research Paper R5, Washington, DC).

Nakicenovic, N. (1986), *Patterns of Change: Technological Substitution and Long Waves in the United States*, Working Paper WP–86–13 (International Institute for Applied Systems Analysis, Laxenburg, Austria).

Nakicenovic, N. and Marchetti, C. (1979) *The Dynamics of Energy Systems and the Logistic Substitution Model*, Research Report RR–79–13 (International Institute for Applied Systems Analysis, Laxenburg, Austria).

Nordhaus, W.P. (1973), The allocation of energy resources, *Brookings Papers on Economic Activity* **3**.

Pearl, R. (1922), The population problem, *Geographical Review* (Oct.): 638.

Shannon, C. E. (1948), A mathematical theory of communication, *Bell System Technical Journal* **27**.

Shannon, C.E. (1951), Prediction and entropy of printed English, *Bell System Technical Journal* **30**: 50–64.

Stiglitz, J. (1974), Growth with exhaustible natural resources: Efficient and optimal growth paths, *Review of Economic Studies.*

Stiglitz, J. (1979), A neoclassical analysis of the economics of natural resources, in Smith, V.K. [ed], *Scarcity and Growth Reconsidered* (Resources for the Future, Washington, DC).

Takayama, A. (1974), *Mathematical Economics* (Dryden Press, Hinsdale, IL).

Tietenberg, T. (1984), *Environmental and Natural Resource Economics* (Scott, Foresman and Company, Glenview, IL).

Tribus, M. and McIrvine, E.C. (1971), Energy and information, *Scientific American* **225** (3): 179–188.

Von Neumann, J. (1945), A model of general economic equilibrium, *Review of Economic Studies* **13**: 1–9.

Von Neumann, J. and Morgenstern, O. (1944), *Theory of Games and Economic Behavior* (Princeton University Press, Princeton, NJ).