Bengt-Christer Ysander

Truth and Meaning in Economics

– Selected Essays on Economic Theory and Policy

IUI The Industrial Institute for Economic and Social Research

The Industrial Institute for Economic and Social Research

is an independent non-profit research institution, founded in 1939 by the Swedish Employers' Confederation and the Federation of Swedish Industries.

Objectives

To carry out research into economic and social conditions of importance for industrial development in Sweden.

Activities

The greater part of the Institute's work is devoted to long-term problems, especially to long-term changes in the structure of the Swedish economy particularly within manufacturing industry.

Board

Address

Industriens Utredningsinstitut
Box 5501, S-114 85 Stockholm, Sweden
Tel. 08-783 80 00

Truth and Meaning in Economics

The Industrial Institute for Economic and Social Research

Truth and Meaning in Economics

– Selected Essays on Economic Theory and Policy

Bengt-Christer Ysander

Editors: Ragnar Bentzel and Gunnar Eliasson

Distribution: Almqvist & Wiksell International, Stockholm

Preface

Bengt-Christer Ysander has had a varied career in many locations. The largest share by far of his academic writing took place during his eight years (1978–1985) at IUI. We–a current and a former president of IUI–find it appropriate to put together a volume of selected writings of Bengt-Christer Ysander for his 60th birthday.

Only articles originally authored in English are included in this volume. Nevertheless, the selection illustrates the broad range of Ysander's academic writing, his originality and his early attention to problems that have only recently been addressed by economists. This collection also provides an opportunity to bring to the attention of a wider audience some of Ysander's original research which has appeared previously only in relatively unknown and not widely distributed and read publications.

Bengt-Christer Ysander's writings cover a broad range of topics, including highly abstract reasoning on the nature of truth in the social sciences (Chapter I), an early theoretical piece from 1981 on non-linear systems dynamics ("chaos") in Chapter II, an insightful essay on the nature of human capital (Chapter III), a sophisticated paper on macroeconomic modeling (Chapter IV), and some very down-to-earth articles on public sector growth and policy making (Chapters VI and VII).

As those of us know who have had the privilege of a long and stimulating association with Bengt-Christer Ysander, his writing is characterized by the same qualities as his personality: keen perception, sophistication, a high degree of technical competence–yet his argumentation is always clear, penetrating, and relevant. We are happy that the readers of this volume now have the opportunity to discover these truths for themselves.

Uppsala and Stockholm, April 1991

Ragnar Bentzel Gunnar Eliasson

Contents

	Page
CHAPTER	
Truth and Meaning of Economic Postulates	11
in Hegeland, H. (ed.), *Money, Growth, and Methodology, and other Essays in Economics,* in honor of Johan Åkerman, Gleerup, Lund 1961.	
CHAPTER II	
Taxes and Market Stability	23
in Eliasson, G. and Södersten, J. (eds.), *Business Taxation, Finance and Firm Behavior. IUI Conference Volumes, 1981:1,* Stockholm.	
CHAPTER III	
Homogeneity in Education. A comment on the economic theories of education	65
in Bowman, M.J., Sohlman, Å. and Ysander, B.-C., *Learning and Earning. Three essays in the economics* of *education,* National Board of Universities and Colleges, Liber Tryck, Stockholm 1978.	
CHAPTER IV	
Structural Change as an Equilibrium or Disequilibrium Process – An Introduction	131
in Ysander, B.-C. (ed.), *Two Models of an Open Economy,* IUI, Stockholm 1986.	
CHAPTER V	
Measuring Energy Substitution – An Introduction	147
in Ysander, B.-C. (ed.), *Energy in Swedish Manufacturing,* IUI, Stockholm 1983.	
CHAPTER VI	
Local Government and Economic Growth	165
in *IUI 40 years 1939–1979. The Firms in the Market Economy,* IUI Yearbook 1979/1980, Stockholm.	
CHAPTER VII	
Public Policy Evaluation in Sweden	173
in *The Economics of Institutions and Markets,* IUI Yearbook 1986–1987, Stockholm.	
Selected Bibliography	185

CHAPTER I

Truth and Meaning of Economic Postulates

"The nature and significance of economic postulates" has been a central point of methodological controverse for more than a hundred years. This paper will introduce a new way of analyzing the role of postulates in the construction of economic theories suggested by recent findings in the philosophy of science. The methodological discussion was formerly primarily concerned with the question whether the common assumptions of economic theory were true or not. In the last decade the discussion has gained in sophistication and has dealt with the question whether truth in the postulates is necessary and relevant to the validity of the theory. We are going to suggest that a great many of the controversial postulates do not have even an empirical meaning except through their consequences. This anlysis eliminates the thruth problem without compromising the truth concept in the pragmatic fashion. Instead of admitting the distasteful notion of "half-truth" into economic postulates it merely concedes the existence of theoretical elements with only a partial empirical meaning. This means that instead of arguing about the "realism" of statements about psychological variables like economic anticipations and maximization normes or limiting concepts like perfect competition and horisontal demand curves, we agree to treating these terms as theoretical variables on a par with terms like "electron" and "instantaneous velocity" in physical theory. We think that this approach will help to clear up a good deal of misunderstanding involved in the old time controversy and bring out some common problems in physical and social science. In the first section of the paper we will recall the main points of the recent discussion on the role of economic postulates and then try to narrow the central problem by picking out the sort of economic assumptions that give rise to difficulties and contain the seeds of controversy. The second section will deal with the status and empirical meaning of postulates and will develop the necessary instruments for this discussion by recounting some recent results in the philosophy of science. The main task of the section is to familiarize the reader with the concept of "theoretical variables". In the third and last section this concept is made use of in the analysis of some appropriate examples of postulates in economics. Since our way of analysing bears an apparent resemblance to the pragmatic analysis so brilliantly put forward in the writings of professor Milton Friedman, we have deliberately chosen to demonstrate our divergent point of view by reanalysing some of professor Friedman's own examples.

I. TRUTH AND HALF-TRUTH IN ECONOMIC POSTULATES

A. The modern discussion concerning the postulates of economic theory was touched off by the pointed summing up of orthodox economics given in Lionel Robbins: "Nature and Significance". His position is stated in the now classical words – "The propositions of economic theory, like all scientific theory, are obviously deductions from a series of postulates. And the chief of these postulates are all assumptions involving in some way simple and indisputable facts of experience relating to the way in which the scarcity of goods which is the subject matter of our science shows itself in the world of reality . . . We do not need controlled experiences to establish their validity: they are so much the stuff of our everyday-experience that they have only to be stated to be recognized as obvious". (Robbins, 1937 pp. 78 ff.). While the postulates of physical science may be both abstract and of disputable validity, the central assumptions of economic theory according to this view are not only true but self-evident. Aside from these main postulates Robbins also mentions the existence of special assumptions stating the concrete circumstances under which the general postulates are applicable: "Being assumptions of a more limited nature based upon the general features of particular situations or types of situations which the theory is to be used to explain". (Robbins, 1937 p. 100)

The next stage was marked by a growing uneasiness about this orthodox complacency. A contrary view was put forward in a most uncomprising manner by Hutchison, who called for a testing of all the central economic assumptions. He wants: "Empirical verification . . . of the fundamental assumption" (Hutchison, 1938 p. 83) and tries to show also that a more realistic theory necessarily will mean "more realistic assumptions" (Hutchinson, 1938 p. 120). At about the same time we find a group of Oxford economists with the same bag of ideas trying to find out by interviews mainly whether the modern business man really recognizes maximization of short run profit as the main goal and norm of behaviour. Most of the answers turned out to be negative; the business men said they wanted a "normal" or "fair" profit, they put their price according to some cost-plus-system, i.e. by adding to the direct cost – as far as marked conditions would permit – a proportional amount covering common costs and a fair margin of profit (Hall, 1939 passim). It has often been pointed out later that the outcome of these interviews does not amount to very much. Expressions like "fair" or "normal" and alibis like "as far as market conditions permit" are too vague to provide a frame work for a theory of business behaviour.

A way of getting around the need for empirical testing of the economical postulates without going back to Robbins' trust in intuition was shown by professor Friedman. In his essay "The Methodology of Positive Economics" Friedman states as his view that economic postulates like other scientific assumptions can only be tested indirectly by way of testing the consequences (Friedman, 1953 pp. 3 ff.). The choice of postulates is not a matter of truth or of empirical validity but can only be decided by reference to the usefulness of the whole theory. According to this pragmatic analysis a system of postulates is a good one if it works well by making possible true predictions and satisfactory explanations. This really means that you abandon the concept of truth as irrelevant in dealing with the postulates of economic theory. Friedman explicitly states that economic assumptions necessarily include both outright falsehoods and "half-truths". The pragmatic approach also closes the door to any further analysis of the structure of meaning in economic theories and the role of assumptions in the construction of new hypothesis. A rather dangerous byproduct of Friedman's pragmatism is his belief that a theory should gain in credibility by having been long upheld by custom and traditional dogma (Friedman, 1933 p. 23).

A new phase in the discussion seemed to be signalled by Koopmans' essay on "The Construction of Economic Knowledge" (Koopmans, 1957 pp. 127 ff.). Taken at face-value Koopmans' essay looks like a return to the standpoint of Hutchison according to which "direct and indirect testing of postulates" remains the most essential task of economists. Further reading, however, makes it more and more doubtful whether Koopmans is really concerned with our problem, i.e. the testability and truth of economic assumptions. The main part of the work deals with possible improvements in economic postulates, e.g. by making the theory of production account for factors of lumpiness and localisation and by including not only risk but also uncertainty as necessary parts in the theory of markets. Whether these improvements can be put to empirical tests is not discussed. Koopmans' call for "more realistic assumptions" turns out to be correspondingly vague in defining the target. Sometimes he writes about the need for better factual estimates and more data – e.g. more detailed technical coefficients in input–output models and more complex production-functions. At other times Koopmans seems to aim at assumptions that "work better", as Friedman would put it, e.g. in the discussion of alternative models of decision.

The analysis of professor Machlup (Machlup, 1955–56 passim) seems at times to underwrite the views developed in this paper especially when touching on "the heuristic postulates and idealized assumptions in abstract models of interdependent constructs useful in the explanation and prediction of observable phenomena" (Machlup, 1956 p. 486). In his conclusions, however, Machlup takes leave of empiricism altogether and is ready to invest even "pure fiction" with empirical meaning on pragmatic grounds and to allow deductive theories alibis in the form of non-testable "Assumed Conditions" against any try to prove them false. – A much closer correspondence with our views is to be found in Mr. Rotwein's discussion of Friedman's essay (Rotwein, 1959 pp. 570 ff.).

B. Even from this short survey of the discussion concerning truth and validity in economic postulates one thing seems evident. Some of the differences of opinion seem to be caused by the ambigousness of the term postulate (Hegeland, 1960 pp. 130 ff.). Not unfrequently one gets the feeling that the various economists are using the same name for quite different kinds of economic postulates. And yet every economist is familiar with at least four distinct kinds of postulates. Even a relatively simple theory like the Cournot-theory of duopoly contains examples of these four kinds of assumptions:

1. An assumption concerning the existence of a demand function. This assumption is an example of that big group of economic postulates which deal with the existence and stability of empirical correlations.

2. An assumption that variations in cost are negligable. This is a special instance of a ceteris-paribus clause and can be taken to represent all those assumptions which specify the circumstances under which a theory is applicable.

3. An assumption of maximization of short run profit. This assumption is typical of the main body of fundamental economic postulates which deal with subjective goals, norms and forms of reasoning.

4. An assumption of economic anticipations. This sort of postulate is fundamental in all dynamic theories. (Cournot, 1938 pp. 88 ff.)

Much confusion and unnecessary disagreement could probably be avoided by clearly distinguishing between postulates of type 3 and 4 involving subjective variables which we do not know how to test, and on the other hand postulates of type 1 concerning empirical correlations the testing of which presents no fundamental problem. Koopmans' discussion and suggested reforms seem to be primarily concerned with assumptions of type 1. The frequent paradoxes in the reasoning of Hutchison and especially of Hall and Hitch seem to be partly founded on a mix-up between assumptions of type 1 and assumptions of type 3 and 4. Our own discussion like the analysis of professor Friedman deals primarily with the latter kind of postulates.

C. We are trying in this paper to analyse those economic postulates which do not seem to be directly amenable to verification. We have already noted an important subclass of these, *viz.* the postulates involving subjective terms. What we want to find out is whether we have to go along with professor Friedman and allow falsehoods and half-truths in these fundamental assumptions. Before going into this problem, however, we want to limit our discussion by specifying certain kinds of "inaccuracies" which fall outside our field of inquiry. There are at least two kinds of half-truths that need not detain us in this context:

1. Most "laws" and correlations in the social sciences are intended to be and must be tested as stochastic distribution laws although the stochastic formulation is often left out for convenience sake and regarded as a tacit assumption. This means of course that these propositions can never be verified in their exact sense nor can they ever be definitely accepted as truth, since any inductive proof according to some statistical criterium can only result in a preliminary acceptance due to eventual correction as further evidence comes along.

2. There is always a certain vagueness in the rules which govern the application or interpretation of the terms of a theory into the world of reality. There can never be an exact agreement on what an economic term should measure or how it should be measured.

The "inaccuracies" or "half-truths" which concern us here stem from the presence in many economic postulates of terms for which there is no easy empirical interpretation at all. Among these terms are limiting concepts like perfect competition or perfectly elastic demand and psychological variables like anticipations and norms of behaviour.

II. DO THE POSTULATES NECESSARILY "MEAN" ANYTHING?

In this section we are going to turn our interest from the question of truth to the question of empirical meaning in postulates. Instead of arguing about the possible inaccuracy of fundamental propositions we are going to suggest that many postulates have only a partial empirical interpretation, involving terms which function as instruments of theory construction rather than as symbols of concrete things. To clear the way for this analysis we have to preface the discussion with some remarks on the "meaning of meaning", i.e. what is really meant by the term "empirical meaning".

The search for a criterion by which it is possible to delimit propositions of empirical science from logical truth, methaphysical speculation and pure nonsense, has continued ever since science first tried to establish itself as independent of church and traditional philosophy. The modern discussion was started under the impact of the accelerated advance of natural science at the end of the last century. The best known of the suggested criteria is the so called verifiability-theory which equates empirical meaning with the possibility of verification (Carnap, 1936 passim). Also wellknown are the snags and difficulties which this theory has run into and which has lead many people to consider the whole attempt a blind alley and foredoomed failure. Not even radical modifications of the principle seem able to make it into a workable definition of the body of scientific propositions (Hempel, 1951 passim).

The usual way to meet these difficulties has been to supplement or substitute the verifiability principle with some conditions concerning the possibility of "translating" the terms of a theory into the language of concrete observations. To "translate" means then to be able to define at least indirectly or conditionally all those words in a scientific proposition which are not logical constants or variables in terms of observable properties and things. A criterion on these lines turns out, however, to be too narrow to be useful. Several fundamental terms of science would not be admittable according to such a criterion. We would, e.g., not be able to express length as irrational numbers; the sentence: "Since both sides in this rightangled triangel are equal to 1 meter the hypotenuse is equal to $\sqrt{2}$ meter", would have to be rejected as without empirical meaning since we do not know any method of measuring $\sqrt{2}$ meter exactly. The same thing is true of all so called "limiting" or "ideal" concepts like instantaneous velocity, absolute vacuum, perfect competition and so on. A comparable difficulty attends the empirical interpretation of abstract terms like electrons, waves of energy, preference functions and many subjective terms like anticipation or goal direction. These uninterpreted terms usually occur in theoretical propositions with empirical consequences. In this way the abstract terms may be said to gain a partial empirical interpretation. This also means, however, that we can not talk of the empirical meaning of a proposition without defining this meaning in relation to the total theoretical system. What a given proposition "means" empirically, obviously depends on two things: Firstly it depends on the logical system which regulates the construction of the sentence and decides what is implied by the sentence, and, secondly it depends on the existence of other established propositions which decide the operational possibility of testing.

In this fashion one has been forced to recognize the necessity and usefulness of "theoretical variables", i.e. terms and parts of theories which can only be partially and indirectly interpreted but still fulfill a useful function in the structure of our knowledge and whose meaning is derived from their observable consequences. Obviously these theoretical variables are especially abundant in the postulates of a theory. Their presence constitutes a challenge to every empiricist. This challenge has of late years called forth three main kinds of suggested solutions, which we can call the functional, the pragmatic and the instrumentalistic approach. (Scheffler, 1957 passim)

1. The functional approach starts with a refusal to acknowledge any principal necessity for theoretical variables in scientific systems. Their essential function is said to be to simplify and shorten the formulation of knowledge. (Nagel, 1956 pp. 103 ff.) Theoretical variables are considered a sort of short hand notions which could principally be eliminated. This idea is backed up by the proof that it is always possible to take the observation terms in a theory and construct out of these a theory which is functionally equivalent to the original theory, i.e. which implies the same consequences. (Craig, 1953 p. 31) Usually this would, however, be immensely unpractical since it would result in the original postulates being replaced by an infinite number of postulates in the equivalent theory. A more important argument against the functional approach is to be derived from the fact that the use of the theory for prediction and explanation involves not only deduction but also reasonings of an inductive kind, which are not possible without the use of theoretical variables. One empirical observa-

tion often adds probability to a theoretical postulate which implies another observation sentence. These internal relations between the observation sentences cannot be expressed without using the theoretical postulate. – The criticism can be taken further by demonstrating the positive function of theoretical variables in the construction of knowledge. The theoretical terms are not only short hand notions but are often the means of tracing unknown and unsuspected empirical variables and correlations, whose existence are not implied directly by the observations. (Braithwaite, 1953 pp. 50–87) Even if we could analyse our existing knowledge without bringing in theoretical variables, which is still somewhat doubtful, we certainly can not dispense with them when it comes to the building up of new theories and the incorporation of new knowledge. – Some of the more extreme formulations in Hutchinson seem close to the functional approach.

2. The pragmatic approach means that you have definitely abandoned the idea of a criterion of empirical meaning, and thus considers the question of validity of postulates to be both unnecessary and irrelevant. According to this view you can not analyse the theory by parts but can only judge its value by considering the usefulness of the total theory, i.e. how far it seems to deliver good predictions and satisfactory explanations (Carnap, 1950 passim). No theoretical proposition is a priori nonsense or false. Instead of looking for validity and empirical meaning in the postulates you will have to base your choice on more vague criteria like simplicity, consistence with established theories and so on (Quine, 1953 passim). For a scientist brought up in the empiricist faith the pragmatic analysis seems a rather desperate council. – Milton Friedman represents among the economists a mature example of radical pragmatism. He has been brought to this standpoint not by discussion of theoretical variables in the economic postulates but on the contrary by totally neglecting this problem and concentrating on the problem of validity. The question of whether or not an economic postulate is empirically valid would seem, however, to depend on the possibility of giving the assumptions an empirical interpretation.

3. The instrumentalistic approach accepts the theoretical variables as necessary instruments for the constructions of theories and concentrates the efforts in trying to explain their function and formulate the rules governing their use (Scheffler, 1957 passim). This is the approach we would consider most fruitful.

III. THE FUNCTION OF THEORETICAL POSTULATES IN ECONOMIC THEORY

In this last section we will try to apply the instrumentalistic approach to the analysis of some of the simple examples of postulates that Friedman uses in his essay. We want to show that instead of just neglecting the question of truth and allowing downright false-hoods in the postulates like the pragmatist does, the postulates can be analysed as containing theoretical variables which makes them inaccessible to direct testing and indeed makes nonsense of the whole problem of validity.

Many of these non-empirical terms were originally used to construct "models of calculation" setting out the most rational way to handle economic problems (Åkerman, 1960 pp. 269 ff.). Later these calculi were inserted into causal explanations to account for the action of entrepreneurs, consumers and so on. The non-empirical terms became what the sociologists call "inferred variables". Familiar names like "anticipations" and "behaviour norms" helped to create the illusion that the "models of calculation" gave a substantially true although simplified picture of "what made people tick". Further reflection has shown, however, that these terms cannot easily be identified with any empirical properties of human behaviour. Nevertheless it is quite possible that they can prove adequate for their purpose and help to provide good predictions. When e.g. you solve the equations describing Cournot's duopoly situation you get "reduced forms" which are amenable to test, involving only the supplies of the two competitors in consecutive periods.

There is, however, one rejoinder which we must expect to meet. To many it may seem both farfetched and fantastic to try to "make nonsense out of perfectly normal and familiar things like perfect competition and economic anticipations". Even in physics the natural aversion to abstract principles has shown itself in various attempts to present intuitive models of, e.g., nuclear theory or wave propagation (Cassirer, 1957 pp. 459 ff.). One thinks about the atom as a compact corpuscular body and visualizes the light or soundwave as vibrations in a string. In 19th century physics it was even usual sometimes to present several rival mechanical models of a theory – as done, e.g., by Maxwell in his "Electricity and Magnetism". The scientific development during the last century has clearly shown, however, that these models are pedagogical aids and heuristic devises and nothing else. Within the social sciences the resistance to accept abstract concepts without intuitive aid will undoubtedly be proportionally greater, since the objects of inquiry are "matters of common knowledge". The deciding criterion must, however, still be the way a concept functions and the possibility of a direct interpretation of it. In this respect there is no principal difference between the two limiting concepts: instantaneous velocity and infinitely elastic demand. And though we may think that we have "introspective acquaintance" with economic anticipations and behaviour norms, when we start testing by interviews or statistical estimation we invariably find eventually that what we are testing are the consequences rather than propositions about the original notions. What we are proposing here is simply that we draw the obvious consequence and regard these original notions as only indirectly interpreted. This does not mean, of course, that we should suddenly stop thinking in terms of our favourite models and use some mathematical symbols instead of talking of anticipations. Our suggestion concerns analysis of theory – not habits of thought.

Friedman gives in fact several good examples of the use of models to aid the understanding of abstract principles. In one passage he considers the problem of predicting how a good billiard player will play his shots. "It seems not at all unreasonable that excellent predictions would be yielded by the hypothesis that the billiard player made his shots *as if* he knew the complicated mathematic formulas that would give the optimum directions of travel, could estimate accurately by eye the angles and so on, describing the location of the balls, could make lightening calculations from the formulas, and could then make the balls travel in the direction indicated by the formulas (Friedman, 1953 p. 21)." This is indeed a good example of what we have been talking about. The assumption we do make concerns the existence of a certain function of stochastical character from which consequences can be drawn as to the probable direction of the billiard player's shots. We have as yet no possibility of directly interpreting the various parts of this function, but we might familiarize ourselves with the function by thinking of it in terms of an intellectual giant. – The same reasoning applies to another example used by Friedman, where he refers to the old observation that leaves tend to be distributed around the tree *as if* every leaf deliberately tried to maximize the amount of sunlight. Friedman rightly points out that the way we think about this distribution theory is irrelevant. The only important thing is the possible existence of some abstract function governing the way the leaves tend to assemble.

He draws the conclusion that the obvious untruth of his antropomorphic model of the function does not matter, while we prefer to draw the equally obvious conclusion that we have no complete interpretation of such a function (Friedman, 1953 pp. 19–20; Koopmans, 1957 p. 139).

The above reasoning can easily be transferred to Friedman's examples from economic theory. After surveying the discussion about the realism of using marginal analysis, Friedman concludes that no assumptions about marginal analysis are completely true but that this does not matter as long as the theory of markets works. We would consider an assumption about marginal curves as only partially interpreted. There is no possible way to get a satisfactory test either on the existence of a certain marginal curve or on the assumption that business men use these curves in a certain way. This means that the assumptions of marginal analysis are to be treated as involving theoretical variables, i.e. as neither true nor false. – A parallel discussion can be followed through concerning assumptions about business men trying to maximize their short run profit. In this case also, Friedman's answer is an irreverent shrug of his shoulders. To us it seems preferable to treat the maximized function as a theoretical variable which receives a partial interpretation by our observations on price behaviour and business planning.

Friedman also deals with some examples of limiting concepts like perfect competition and infinitely elastic demand. He points out that these concepts can never be identified in the empirical world but can be considered "more or less satisfactory" descriptions. In some cases they are good enough while in many other situations they can not be used for producing good predictions and explanations. To our mind the crux of the matter is not that these concepts are sometimes "accurate enough" but that without these limiting concepts, which can not themselves be identified anywhere, we could never hope to measure, e.g., degrees of competition or to get testable propositions about price movements and profit distribution.

•

Our main intention with this paper has been to suggest a somewhat new approach to the analysis of economic theory construction. We have tried to show that abstract, ideal or subjective concepts in economic postulates should be treated as theoretical variables with only a partial empirical interpretation. This does not necessarily imply any denial of an ultimate reality behind these concepts (Rozeboom, 1960 passim). It only restricts the empirical meaning of these terms as used in current economic theory. – Does it really matter then, what we call these concepts? Is the whole problem worth discussing? Our answer to questions like these would be that we hope by this approach to be able to avoid some of the time-consuming controversy over "realism of economic assumptions" and to steer methodological discussion onto the more fruitful problems of when and why theoretical variables are permissible.

REFERENCES

Braithwaite, R., *Scientific Explanation*. Cambridge University Press, 1953.

Carnap R., "Testability and Meaning", *Philosophy of Science*, 4 (1936) and "Empiricism, Semantics and Ontology", *Revue internationale de Philosophie*, 11 (1950).

Cassirer, E., *The Philosophy of Symbolic Forms III*. Oxford University Press, 1957.

Cournot, A., *Recherches sur les Principes Mathématiques de la Théorie des Richesses*. Paris: Marcel Rivière & Cie, 1938.

Craig, W., "On Axiomatizability Within a System, "*Journal of Symbolic Logic*, 18 (1953).

Friedman, M., *Essays in Positive Economics*. University of Chicago Press, 1953.

Hall, R. and Hitch, C.: "Price theory and Business Behaviour", *Oxford Economic Papers*, 2 (1939).

Hegeland, H., "Om möjligheten av dynamisk ekonomisk teori", (On the Possibility of Dynamic Economic Theory), *Statsvetenskaplig tidskrift*, 63 (1960).

Hempel, C., "The Concept of Cognitive Significance", *Proceedings of the American Academy of Arts and Sciences*, 80–1 (1951).

Hutchison, G. W., *The Significance and Basic Postulates of Economic Theory*, London: Macmillan and Co., 1938.

Koopmans, T., *Three Essays on The State of Economic Science*. New York: Mc Graw-Hill Book Company, 1957.

Machlup, F., "The Problem of Verification in Economics" (1955) and "Rejoinder to a Reluctant Ultra-Empiricist" (1956), *Southern Economic Journal*, 22.

Nagel, E., *Logic Without Metaphysics*. Glencoe: The Free Press, 1956.

Quine, W. v., *From a Logic Point of View*. Harvard University Press, 1953.

Robbins, L., *An Essay on the Nature and Significance of Economic Science*. London: Macmillan and Co., 1937.

Rotwein, E., "The Methodology of Positive Economics", *Quarterly Journal of Economics*, 73 (1959).

Rozeboom, W., "Studies in the empiricist theory of Scientific Meaning", *Philosophy of Science*, 27 (1960).

Scheffler, I., "Prospects of a Modest Empiricism", *Review of Metaphysics*, 10 (1957).

Åkerman, J., *Theory of Industrialism*. Lund: Gleerup, 1960.

CHAPTER II

Taxes and Market Stability

Much has been said and written by now about the possible stabilizing effects of public budgets on the effective demand in the total economy. On the following pages we are concerned with a hitherto seldom discussed topic, namely the possible destabilizing effects of taxes and subsidies in individual markets. Particular examples of these possibilities, for example in the labor markets and in the markets for housing, have lately aroused a good deal of public discussion in Sweden, whose world leadership when it comes to taxing ambitions, especially marked in the seventies, makes some of these problems particularly acute. Unfortunately we still lack a well-established analytical framework for dealing with these kinds of stability problems. The modest aim of the following discussion is merely, to point out some dimensions of the problem and to provide some illustrative examples of possible tax-induced instability.

The Changing Role of Taxation

Over the last half-century "taxation"--which in the following I take to include also negative taxes or subsidies--has not only been steeply increased in most market-economies, but has at the same time also changed character. Taxation once used to be dominated by the fiscal aim of financing the provision of certain basic collective goods, mainly the machinery of control--central administration, defense, justice, etc. The means, then, could be kept relatively few and simple--a low income tax with at most a mild form of progression and/or local estate rates. This, as it happens, is still the picture of the public sector often presented in economic equilibrium theory--the provision of collective goods being financed if not by lump-sum taxes then by some proportionate taxation on final goods. There is, then, no need to worry about taxes destabilizing individual markets. Apart from the problems of international adjustment, in a model economy without monetary markets proportionate price increases need not change the stability properties of individual product markets.

The aims and means of taxation today are very different. A drastic illustration of this is provided by Sweden, where the structural change in taxation has probably gone further and faster during postwar years than in any other industrialized market economy.

The provision of collective goods in the narrow definition of the word presented above, plays a steadily decreasing role in the public budgets and is now responsible for less than 15 per cent of total central government expenditure. Apart from social insurance the dominant expenditure items on the public budgets are, now, subsidies of social and private goods. In the national accounts these are classified either as public consumption or transfers depending on how production and distribution are organized.

The ways of financing public expenditure are also becoming more varied and complex. Although taxes on income and wealth, have been sharply increased and made more progressive in the early seventies, they now provide, in Sweden, less than half of central government income and are, to an increasing extent, being complemented by various forms of indirect taxation, including, V.A.T., obligatory social insurance fees and taxes on non-labor factors.

Today's public budgets, therefore, can be best characterized as huge instruments for central price and income regulation. By combining positive and negative taxation with various forms of tax rebates and subsidy rules a highly individualized and differentiated form of taxation can in principle be realized--given the necessary information. With the high general level of taxation--more than 2/3 of private disposable income being channeled through public budgets--the tax effects on individual markets are, in any case, becoming increasingly decisive for price-setting and profitability also in the private production sectors.

The differentiation of means are correlated to--and indeed to a large extent motivatived by--a differentiation of the aims of taxation. The central government's wish to fulfill increasingly differentiated aims concerning industrial and regional policy and income redistribution without undue centralization of market decisions, have put a great strain on the system. In the last few years the shrinking possibility for redistribution in Sweden by way of progressive income taxes has led to an increased use of differentiated price subsidies as a means of redistribution.

There are doubts as to whether we have - or will ever have - sufficiently precise tax instruments, and enough information on how to use them, to match the regulatory ambition of the government. Most tax instruments are still rather blunt in the sense that considerations of fairness and administrative simplicity force us into making tax rules so general that they usually hit rather widely or wildly compared with the aims of tax policy. The complex pattern of taxation and the decentralized handling of various policy areas also make it increasingly more difficult to discern or guess the combined impact of the various horizontal chains of taxation on individual markets and goods.

This raises several important questions concerning efficiency limits to economic control by way of taxation. The one we are going to deal with here is the problem of possible tax-induced market instability. What happens to "normal" price adjustment mechanisms when these are not only transformed by prevailing tax rates but also intercepted by a simultaneous process of tax adjustment with a quite different purpose? How do the "tax links" between different markets affect the stability of interrelated markets? What are the chances of attempted tax adjustments ever converging on the intended allocative or distributional targets?

Market Stability from An Equilibrium Point of View

In looking for an analytical framework for studying tax-induced market instability you are faced with two main alternatives. You can plunge directly into a disequilibrium scenario, which means paying the price of not being able to generalize and of not necessarily ever being in the neighborhood of equilibrium.

The other and more traditional way of studying stability problems is by looking at them from the point of view of an equilibrium position. The question will then roughly be the following: given that the agents behave as if they were constantly in an equilibrium and that the adjustment process follows some simple prescribed rules, what are the conditions for convergence? The results you attain this way are mostly of a rather formal and general nature, but may still provide some leads as to how to structure our approach to the problem of tax-induced instability.

The usual stability analysis aims at determining sufficient conditions under which a system of market price adjustments, each being a monotonic function of excess demand, will converge.[1] The results of these studies are by now well known (cf., for example, Karlin (1959), Lancaster (1968), Arrow-Hahn (1971)). To make sure of convergence three types of conditions are usually needed. One type of condition guarantees that the agents are willing to accept disequilibrium prices as if they stemmed from a final equilibrium (cf., "Walras' law"). A second type of condition--for discrete-time adjustments--is needed to ensure that the rate and/or stepsize of adjustment is not so big that you over-shoot the equilibrium target by too much.

---

[1] See Appendix, note I.

Finally you need some condition concerning the links between the adjustment in different markets to make sure that solving excess demand problems in one area does not inflate the same problems in other markets by too much.

This last condition can take many technical forms --"gross substitution", "aggregate revealed preference", "diagonal dominance", etc.-- all of which, unfortunately, appear rather restrictive and difficult to make intuitively plausible.

These conditions are suggestive when transplanted to our special problem of tax-induced instability. When agents become conscious of prices being to a large extent determined in government offices, they may be less willing to accept them as given data to which they passively adjust. The varying "tax multiplier" on price in different markets could increase the risk for excessive, destabilizing adjustment steps in some markets.

Taxes and tax adjustments tend to provide direct links between adjustments in different markets. The risk would consequently increase that an adjustment in one market might counteract overall stability by disrupting other markets.

There are other limitations of existing economic stability analysis apart from the restrictive conditions used. It tells us, in fact, little or nothing about those stability properties of the economic system that we are often most interested in when dealing with real-life economies.[1] One

---

[1] See Appendix, notes II and III.

such property, for example, is stability in the sense that prices (and volumes) originating from a point within a region will never move outside given boundaries. Another question has to do with the possibility of prices converging to an equilibrium "close" to the original one, after a shift in some coefficient. In as far as taxes tend to change even the behavioral structure of an economic system these stability questions are very pertinent and will be raised again later on in connection with some of the illustrative examples quoted.

The problem with which we are concerned here--simultaneous price and tax adjustment in individual markets--can obviously be treated as an extension of the traditional market stability problem. The stability problem of decentralized policy, without involving simultaneous price adjustment, has been discussed by inter alia Mundell (1962) and Cooper (1967). They were concerned with the risks of instability with a decentralized policy arising from the inability of individual authorities to foresee and take into account the effects of policy instruments on markets or areas outside their own field of responsibility. The question of what happens if you combine the two problems--superimposing a tax adjustment on a market price adjustment--has, however, not been treated in economic literature, as far as we know. We hope the examples presented below will suffice to show that further work in this direction could be worthwhile and relevant to economic policy.

Tax-induced Instability in A Single Market

Let us start by looking at a general and very simple case --price- and tax-adjustment in continuous time in a single market. The "tax coefficient", T, is supposed to be defined in terms of the producer price, P. The product, TP, gives the demand price. The producer price is supposed to adjust in a simple way, changing in proportion to excess demand, while the tax rate is adjusted proportionate to some other function of market conditions. A straightforward tax target--relatively innocuous from a stability point of view--would be the volume of demand. The aim of the tax authorities could then simply be to make demand, d, adjust to a pre-set value d*. The purpose of such a tax target could be, for example, to limit the effect of environmental damage or some other collective externality or to keep down consumtion of some noxious commodity. Denoting the supply function by s(P) we would then have the following system:

$$P = \alpha E = \alpha\big(d(TP) - s(P)\big) \qquad (1)$$

$$P,\ T,\ \alpha,\ \lambda > 0$$

$$\dot{T} = \lambda G = \lambda\big(d(TP) - d^*\big) \qquad (2)$$

If we assume stability in the Liapunov sense, local asymptotic stability or resilience[1] is a necessary condition for global stability. With this assumption we can discover possibilities of

[1] See Appendix, note I.

global instability by simply looking at local properties.[1]

If we assume E and G to be continuous functions and P*,T* to be an equilibrium point, we can use a linear approximation around this equilibrium;

$$E = E_p^* p + E_\tau^* \tau \tag{3}$$

$$G = G_p^* p + G_\tau^* \tau \tag{4}$$

where E_p^*, G_p^*, E_τ^* and G_τ^* denote the first partial derivatives of E and G with respect to P and T at the equilibrium, and p, τ stand for (P-P*) and (T-T*).

The linear adjustment system can then be written in vector form as:

$$(\dot{p}, \dot{\tau}) = A(p, \tau)^2 \tag{5}$$

where A is the matrix

---

[1] It should perhaps be emphasized that what we are, then, conditionally proving is only that the system will not tend to work back all the way to the equilibrium. To prove unconditionally that the system is unstable in the sense of Liapunov, that it will eventually tend to cross any preset boundary, would require, for example, the use of one of Liapunov's own instability theorems and would in the discussed examples be a difficult --and often impossible-- task.

[2] A tax adjustment similar from a stability point of view is implied by any progressive taxation of the supply price. This can be seen, for example, by writing the progressive rate as $T = \lambda P$ which gives $\dot{\tau} = \lambda \dot{p}$.

$$A = \begin{matrix} \alpha E^*_p & \alpha E^*_\tau \\ \lambda G^*_p & \lambda G^*_\tau \end{matrix} \qquad (6)$$

It may facilitate the understanding of the adjustment process if we rewrite (5) in terms of the slope o0 the demand curve, $\partial d^*/\partial(TP)$, and supply curve, $\partial s^*/\partial P$, respectively:

$$\dot{p} = \alpha(pT^*+\tau P^*)\left(\frac{\partial d}{\partial (TP)}\right)^* - \alpha p\left(\frac{\partial s}{\partial P}\right)^* = \alpha\Delta^*(d-s) \qquad (7)$$

$$\dot{\tau} = \lambda(pT^* + \tau P^*)\left(\frac{\partial d}{\partial (TP)}\right)^* = \lambda\Delta^* d, \qquad (8)$$

where Δ is used to denote the differential. In comparison with a market situation without tax, two changes have occurred in the adjustment. The demand differential is now a function of two kinds of divergences instead of just one--in the producer's price and in the tax coefficient. Secondly, beside the price adjustment we now have the tax adjustment being proportionate to the change in demand as well.

The system (5) is a first order homogeneous linear vector differential equation. It will converge--showing local asymptotic stability--if and only if all roots of A have negative real parts.[1]

---

[1] For a survey of the "mathematics of stability" cf. La Salle-Lefschetz (1961) and Murata (1977). See also Appendix, notes I-II.

The two roots, x_i, of A are:

$$x_i = \frac{\alpha E_p + \lambda G_\tau}{2} \pm \sqrt{\left(\frac{\alpha E_p + \lambda G_\tau}{2}\right)^2 - \alpha\lambda(E_p G_\tau - G_p E_\tau)} = \quad (9)$$

$$= \frac{\left[(\alpha T^* + \lambda P^*)\frac{\partial d}{\partial (TP)} - \alpha \frac{\partial s}{\partial P}\right]}{2} \pm$$

$$\pm \sqrt{\left[\frac{\left[(\alpha T^* + \lambda P^*)\frac{\partial d}{\partial (TP)} - \alpha \frac{\partial s}{\partial P}\right]}{2}\right]^2 - \alpha\lambda\left(-P^* \frac{\partial d}{\partial (TP)} \frac{\partial s}{\partial P}\right)} = \quad (10)$$

$$= a \pm \sqrt{a^2 - b}. \quad (11)$$

A closer inspection reveals that $a^2 > b$, i.e., the roots are real. No oscillatory price movements will occur owing to the fact that tax adjustment, as defined, follows and reinforces the price adjustment.

Given this, the convergence condition can be written as:

$$a^2 > b \rightarrow \left[\left(a \pm \sqrt{a^2 - b}\right) < 0 \equiv \begin{cases} a < 0 \\ b > 0 \end{cases}\right] \quad (12)$$

Written out in terms of the slopes of the demand and supply curves (12) acquires the following meaning:

$$\frac{\partial d}{\partial (TP)} < 0, \; \frac{\partial s}{\partial P} > 0 \quad (13)$$

This convergence condition should be compared with the condition for stability in the Walrasian sense in a market with only price adjustment:

$$\frac{\partial d}{\partial (TP)} < \frac{\partial s}{\partial P} \tag{14}$$

In the "normal" case with a negatively sloping demand curve and a positive slope of the supply curve, we will have local stability both with and without tax adjustment. However, with supply price decreasing with scale, i.e. the supply curve having a negative slope,--and with the case of demand increasing with price--the risks of instability differ.

Without tax, the price will be instable only if the negative slope of the supply curve is less steep than that of the demand curve. This traditional condition for stability means that the convergent price change via the demand term should in absolute terms dominate an eventual counteracting supply term.

With the tax being determined as in (2), any negatively sloping supply curve will, however, make system (5) instable. This can be intuitively understood from the expressions (7) and (8). We see that divergences in demand price ($pT^* + \tau P^*$) determine the tax change, and also affect the change in the producer's price. The tax in other words, acting as a wedge between supply and demand prices, keeps the demand price from diverging too fast, which in turn makes it possible for the supply price to outrun the demand price.

Without taxes this cannot happen even when supply tends to decrease slightly with price. Suppose supplies are too big, with supply prices being too low. This in itself will tend to lower the price further. Demand, however, will act in the opposite, stabilizing direction. Being more price-sensitive, it will dominate. Introducing a tax wedge means that the demand price can be controlled by way of increased taxation allowing the supply price to slide further without being effectively checked by a demand expansion, etc. The tax has made both prices instable.

Other tax targets may, however, introduce new and potentially larger risks of instability. Local government price subsidies for utilities, housing, etc., in Sweden seem to aim at keeping the household expenditures for these "necessities" constant relative to household income. Let us assume prices to be expressed in some representative numeraire and neglect income changes. This tax target would then mean that current expenditure on the item in question has to be adjusted to some prescribed amount M. In a wider political interpretation this tax rule could be thought of as implying that political decision-makers allocate the subsidies to the big expenditure items so as to maximize appreciation and votes. With this interpretation the rule approximates subsidizing policies within a wide range of state and local areas, from adult education and recreational activities to fringe services on health and old-age care. Keeping the denotations as above, the adjustment system can be written as:

$$\dot{P} = \alpha E = \alpha(d-s) \qquad (15)$$

$$P, T, \alpha, \lambda > 0$$

$$\dot{T} = \lambda G = \lambda(M - PTd). \qquad (16)$$

Using the same reasoning as before, we find that the real parts of the corresponding matrix roots have to be negative for the adjustment system to converge.

The matrix roots are:

$$x_i = \frac{1}{2}\left[\alpha T^*\frac{\partial d}{\partial (TP)} - \alpha\frac{\partial s}{\partial P} - \lambda P^* d(1+e_p)\right] \pm \sqrt{\left(\frac{1}{2}\left[\alpha T^*\frac{\delta d}{\partial (TP)} - \alpha\frac{\partial s}{\delta P} - \lambda P^* d(1+e_p)\right]\right)^2 - \alpha\lambda P^* d(1+e_p)\frac{\partial s}{\delta P}} = \qquad (17)$$

$$= a \pm \sqrt{a^2 - b}, \qquad (18)$$

where e_p denotes the price elasticity of demand. As before, all derivatives are evaluated in equilibrium.

In this case, complex roots may appear giving rise to oscillatory price movements, which is what we would expect since tax and price adjustment in (15-16) tend to counteract each other.

We thus have the following two possibilities of convergence:

I. Dampened oscillation $\begin{cases} a < 0 \\ b > a_2 \end{cases}$ (19)

II. Straight convergence $\begin{cases} a < 0 \\ a_2 > b > 0 \end{cases}$ (20)

The common necessary conditions for convergence, $a < 0$, $b > 0$, can be derived directly from (17):

$$T^* \frac{\partial d}{\partial (TP)} - \frac{\partial s}{\partial P} < \frac{\lambda}{\alpha} P^* d(1+e_P); \; \frac{\partial s}{\partial P}(1+e_P) > 0 \qquad (21)$$

Let us finally also have a closer look at the condition that differentiates between dampened oscillation (19) and straight convergence (20). We will get oscillatory convergence if:

$$4\alpha^2 T^* \frac{\partial d}{\partial (TP)} \frac{\partial s}{\partial P} > \left[\alpha\left(T \frac{\partial d}{\partial (TP)} + \frac{\partial s}{\partial P}\right) - \lambda P^* d(1+e_P)\right]^2 > 0 \qquad (22)$$

One simple implication of (22) is that:

$$\frac{\partial d}{\partial (TP)} \frac{\partial s}{\partial P} > 0.$$

In other words we will get oscillatory convergence only if the supply or demand curve behaves "abnormally", when we have, for example, a negatively sloping supply curve. If condition (22) is fulfilled, the movement of both the supply price and the tax coefficient will be described by:

$$\left[P(t),\ T(t)\right] = k_1\ e^{\rho t}\left[\cos(\nu t+\Phi)r - \sin(\nu t+\Phi)\upsilon\right]$$

$$+\ k_2\ e^{\rho t}\left[\cos(-\nu t+\Phi)r + \sin(-\nu t+\Phi)\upsilon\right] \qquad (23)$$

where $\rho \pm \nu i$ = the roots, $r \pm \upsilon$ = the characteristic vectors associated with the roots, and where both the conjugate constants k_1 and k_2 and the phase constant, Φ, depend on initial conditions.

From (21) we see that with an elastic demand, ($e_P < -1$), and a positive supply curve, subsidies aimed at stabilizing expenditure will introduce instability of price. This is also easy to understand intuitively. While, in the first example, producer prices and the tax coefficient are adjusted in the same direction, thereby slowing down the adjustment of each other we now have a reversed situation. Suppose the producer's price has been set too low. This gives rise to excess demand, moving the supply price upwards. At the same time, however, with elastic demand, expenditures are too big, which means that the tax coefficient moves down. Hence, subsidies grow, counteracting the effect of the producer's price on demand price. This, obviously, leads to a decreasing demand price followed by an increasing supply price, etc.

Taking a gradual increase of both income and of the expenditure target, M, into account does not change this conclusion. A too low supply price then means an increased potential risk of instability compared to a too high supply price. If the subsidy rule is changed to mean that subsidies vary in a fixed proportion to demand, the conclusion is in fact strengthened --holding for an inelastic demand as well. Political expediency may often seem to require the use of such "explosive" subsidy rules. This is illustrated by the Swedish experience in some areas of health and recreation.

The model exemplified above can be generalized to the multi-market case. Without individual specification of the tax rules involved little more can, however, be learned from such a generalization except the important, but obvious, conclusion that none of the ususal sets of sufficient stability conditions retain any credibility when extended to involve also tax adjustment rules.[1]

The step-size of tax adjustment

Real life adjustment is seldom a continuous process. This is true both for price-setting producers and, perhaps even more, for tax authorities.

If we make the realistic assumption that adjustments take place in discrete steps, the size of these steps or the rate of adjustment becomes important for stability.[2]

Since there is, no longer, an immediate feed-back from market reaction to adjustment, you now run the risk of over-shooting your targets. If your "over-correction" is even bigger than the needed correction, the adjustment will obviously become unstable.

This is true already when there is only a price adjustment to deal with. Formulated as a difference equation with $\Delta p(t) = p(t+1)-p(t)$ and $p(t)$ representing the divergence from equilibrium, the price adjustment can be written:

---

[1] Cf. Ysander (1980), where sufficient conditions for the multi-market case are discussed.

[2] In actual life you may, of course, decide independently how often to adjust and how much to adjust. In the analytical example above, however, the time period is taken as given, restricting the possible variation to the rate of adjustment.

$$\Delta p(t) = \alpha E_p^* p(t) \qquad \alpha > 0 \tag{24}$$

By iteration, this can be solved as:

$$p(t+1) = (1 + \alpha E_p^*)^t p(0). \tag{25}$$

The wellknown condition for convergence is:

$$-2<\alpha E_p<0 \text{ (with alternating values for } -2<\alpha E_p<-1) \tag{26}$$

This simply expresses that any "over-correction" must be less than the needed correction. The Walrasian condition for market stability being fulfilled, (26) can be expressed as limits for the rate of adjustment:

$$0 < \alpha < \frac{2}{-\left(\frac{\partial d}{\partial p} - \frac{\partial s}{\partial \bar{p}}\right)} \tag{27}$$

Since any fixed positive tax, T, will increase the step-size of demand-induced adjustment by $(T-1)\alpha$, by definition it follows that even without tax adjustments all proportional market taxes will narrow the safety margins for stable price adjustment.

Let us now take a further step and introduce a tax that is adjusted at the same intervals as price and has the same simple aim as that in our first example above, i.e., to keep demand at a pre-determined value d*. In vector form the adjustment system (neglecting again the asterisks when possible) can be written as:

$$(\Delta p(t),\ \Delta\tau(t)) = A(p(t),\ \tau(t)), \tag{28}$$

where, A, stands for the same matrix as in (6) above and $\Delta(t) = \Delta(t+1) - \tau(t)$ with $\tau(t)$ representing the divergence from an equilibrium tax coefficient, T*. A necessary condition for convergence of a simple difference system of this kind is that:

$$|1 + x_i| < 1; \qquad i=1,\ 2 \tag{29}$$

where x_i is a root of A.

We already know the roots from (9-11) above, and know that they are real. Thus:

$$-2 < a \pm \sqrt{a^2-b} < 0 \tag{30}$$

It was shown in (13) above that the second part of this condition requires that the demand slope be negative and the supply slope positive, i.e., a "normal" market situation. The first part of (30) is the now added restriction on step-size. Given the second part of (30) we can spell out the first part in the following manner:

$$\begin{matrix} a^2 > b \\ a<0,\ b>0 \end{matrix} \rightarrow \left[(a \pm \sqrt{a^2-b} >) -2 \equiv \begin{matrix} a > -2 \\ b > -4(1+a) \end{matrix} \right] \tag{31}$$

The two inequalities to the right in (31) express constraints on the rates of adjustment, α and λ.

$$(\alpha T^* + \lambda P^*) \frac{\partial d}{\partial (TP)} - \alpha \frac{\partial s}{\partial P} > -4 \tag{32}$$

$$\alpha\lambda P^* \frac{\partial d}{\partial (TP)} \frac{\partial s}{\partial P} - 2\left[(\alpha T^*+\lambda P^*) \frac{\partial d}{\partial (TP)} - \alpha \frac{\partial s}{\partial P}\right] < 4 \tag{33}$$

After some reshuffling (32) and (33) yield the following limits --now expressed in terms of the decision variables p and τ-- for the rate of price adjustment, α:

$$\frac{\frac{\partial d}{\partial \tau}}{\frac{2}{\lambda}\left(\frac{\partial d}{\partial p} - \frac{\partial s}{\partial p}\right) - \frac{\partial d}{\partial \tau}\frac{\partial s}{\partial p} + 4} < \alpha < \frac{4 + \lambda\frac{\partial d}{\partial \tau}}{-\left(\frac{\partial d}{\partial p} - \frac{\partial s}{\partial p}\right)} \qquad (34)$$

Comparing (34) with the restriction on α without taxes in (27) (and remembering that the slope of the original demand curve corresponds to $\partial d/\partial$(TP = 1/T*($\partial d/\partial p$)) we see that the introduction of tax means that α is now bounded also from below and that both bounds are functions of the rate of tax adjustment, λ. The right-side inequality shows, for example, that the more price-sensitive demand is, the slower the tax adjustment has to be, given α. Increasing the relative tax adjustment rate, λ/α, will always lead to instability.

Taxes and Structural Stability

Our examples so far have dealt with stability in the usual sense, i.e., we have discussed price developments in a market characterized by given coefficient values.

Of at least equal interest, but more difficult to exemplify formally, is the case where a tax adjustment rule renders the market structurally instable, in the sense that even small changes in the parameters will change the behavior of the system, establishing a quite different set of equilibria or regions of stability.

When we are discussing the stability of an economic system in the face of large quantitative or qualitative changes, say, a big hike in oil prices or drastic changes in the laws governing ownership of firms, the myopic study of local stability properties is seldom of much use. The kind of instability we are then interested in means that we are far from the original equilibrium or the established growth-path. If the initial disturbance concerned the size of an endogenous variable in our model of the economy, we would say that the size of the change had been "out of bounds" for the stability region within which we had, so far, been operating. With the change occurring in an exogenous variable or a behavioral parameter we would, instead, interpret the result as evidence of "structural instability" in the sense that shifts in the parameters can lead to changed stability properties, a new topography for the phase space of the system.[1]

The introduction of taxing procedures on various markets is, in itself, an important change that could modify the structural stability properties of the entire system. Taxes may, moreover, often induce changes in the behavior of the economic actors as well as alter the system's ability to adjust to and absorb other institutional or environmental changes that occur.

The Swedish economy abounds with illustrative examples of tax-adjusted behavior and tax-induced changes in market structure.

---

[1] See Appendix, note III.

High tax rates have, in many cases, led to the establishment of "grey" or "black" markets. Competition from these often modifies behavior in the "official" markets considerably. In the fifties and sixties market structure also tended to change as a result of taxation laws being generally unfavorable to small family businesses. In recent years the combination of complex tax laws, mostly written in nominal terms, and a high rate of inflation have led to huge unintended discrepancies betweeen the tax treatment of various kinds of real and financial investment. Since these discrepancies are quickly discounted in capital values they tend to make the whole economic system increasingly vulnerable to changed expectations of inflation or of tax adjustment.[1]

Any attempt to discuss these structural stability problems in substance would take us far beyond the scope and ambition of this paper. Let us, however, try to clarify the formal stability concepts involved by giving an example from oil price-setting, couched in the same terms of market adjustment as our preceding analysis. The example chosen may fill this function, although it can claim no immediate relevance for policy.

Suppose there are two kinds of oil prices, P_r which is an index of the US producer price of refined oil and, p_o, which stands for an index of the Saudi government's unit charge for crude oil.

---

[1] For an assessment and a discussion of these assymetries and discrepancies in the tax treatment of different kinds of investment cf. Johansson (1978).

It is assumed, here, that the U.S. oil companies try to reduce any eventual gap between their domestic price increase and that of the Saudi government. The Saudis on their side are considered to have an idea of what constitutes a "fair" proportion, r, between the price increase they get and that of the U.S. companies. The price adjustments can then be described by the following:

$$\dot{p}_r = \alpha(p_o - p_r) \tag{35}$$

$$\dot{p}_o = rp_r - p_o. \tag{36}$$

The stability properties of this system obviously depend crucially on r, the Saudi's preset idea of a fair proportion. r=1, for example, means that any point with $p_o=p_r$ is a stable equilibrium. With r>1 no equilibria exist and prices will explode.

The U.S. government now interferes in the game, trying to curb the inflationary impulses of the oil parties by taxing away domestic demand whenever oil price hikes increase. The oil tax rate, τ, expressed as a multiple of p_r, is raised in proportion to the product of both oil prices, although at a decreasing rate. The Saudis now have to take the tax into account in calculating the "fair" proportion. The total adjustment can be written as follows:

$$\dot{p}_r = \alpha(p_o - p_r) \tag{37}$$

$$\dot{p}_o = (r - \tau)p_r - p_o \tag{38}$$

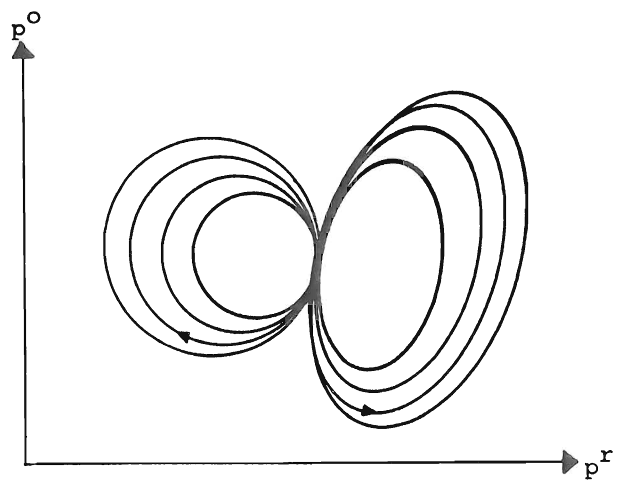
$$\dot{\tau} = p_r p_o - \beta\tau. \tag{39}$$

The behavior of this system is very different. For small values of r the system globally tends to a simple equilibrium. Should the Saudis, unlikely enough, consider it "fair" that the crude price develops much slower than the U.S. domestic price, the effect of the tax may be, in fact, to accelerate the downsliding of both prices towards zero. For a somewhat larger r, there is one stable equilibrium (two, if negative prices are allowed), denoting an equal price increase, with a positive tax to balance off the Saudi's claim for a "fair" price edge.

If r gets even larger--magnifying the Saudi's idea of a fair relation of price--it suddenly leads to a completely new mode of behavior. Wherever the development starts off (excepting some isolated points of equilibrium) it will eventually be drawn into a circular motion of prices and tax. The crude price leads, due to the Saudi's high price ambition, with the U.S. price following. Both are, however, outrun by a fast although decelerating tax change. The high tax then turns the movement downwards, again with the crude price in the lead, followed by tax and U.S. domestic price until the shift in relative oil price is enough to offset the tax and the crude price starts increasing again. The relative oil price will thus vary around 1 while the tax rate moves around (r-1). The development is, however, very sensitive to small differences in the values of the variables. After a certain number of "orbits" (the rotation numbers being a Markov sequence) the system will suddenly branch off into another but similar "orbit", only to return again after a while to the first "orbit", etc. Looking at the system from outside we would observe sudden shifts in the price- and tax-cycles occurring according to a seemingly stochastic schedule. The movement could --projected on the price plane--look like figure A.

This rather "exotic" example[1] illustrates the fact that taxes may not only change the stability properties around equilibria; they can also change the whole nature of equilibria and their structural stability in the face of parameter changes.

Figure A. Alternating price cycles



[1] The quoted model is an instance of the so-called Lorentz model, originally invented to solve a problem in aerodynamics (Lorentz, 1963). It has later been shown to give a good description also of the reversals of Earth's magnetic field over geological times (Ruelle and Takens, 1971). Continued work with this kind of attractor system has been reported by Grümm(1976a-b).

Tax Uncertainty and Market Stability--the Housing Market

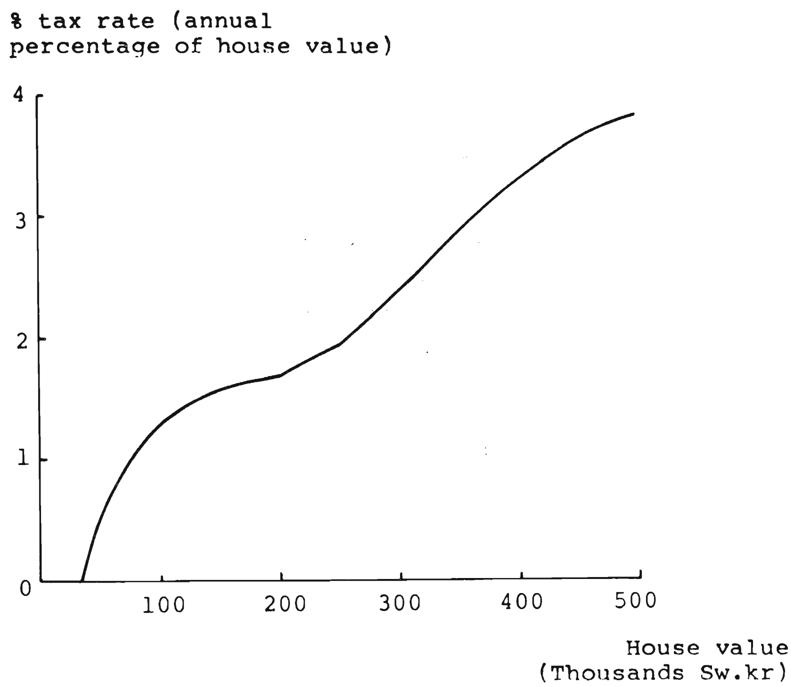
So far we have dealt explicitly only with tax-induced instability under full information. However, if tax adjustment is hard to predict for the parties concerned, the induced uncertainty may give rise to stability problems in the form of highly erratic price movements. A striking example of this is provided by the Swedish market for owner-occupied houses.

Pricing, in this market, is to a large extent determined by the tax authorities. This is done firstly by assessing the taxable value of the property--supposedly at 3/4 of market value--and secondly by applying to this value a progressive scale of imputed taxable income, which is then superimposed on the already steeply progressive income tax. The outcome in many cases is that the owner pays more to the government than to his bank and that what the tax authorities evaluate is in fact the result of previous tax decisions. Especially when tax scales and tax norms are changing rapidly and at an unpredictable rate this can give rise to cyclic price fluctuations and demand instability. In recent years, inflationary gains have dominated houseowners' expectations. Tax instability--which increases with inflation--could soon, however, become a serious problem especially if inflationary expectations also become unstable. A relatively advantageous taxing of capital gains on private houses compares favorably with the level of taxation on more rigidly taxed markets, for example, the stock market and bank deposits. Fluctuating capital gains from private real estate find their way back to other markets and there contribute to intermittent swings in demand.

Let us take a closer look at the way in which unpredictable tax adjustments create instability problems.

The theoretical impact of current property tax rates is shown in Fig. B. We have computed the curve for a recently assessed house whose owner has, on average, a marginal income tax rate of 75 per cent. The curve is "theoretical" in so far as it presupposes that the prescribed assessment norm --3/4 of market value-- is strictly adhered

Figure B. Current Swedish tax rates on owner-occupied houses



to. Actually, this has not been the case in recent years. By systematically lowering the norm for more expensive houses in the most recent assessment (1975), the tax authorities seem to have, to a certain degree, counteracted the effects of progression.

To see what the progressive rates might do to the prices of houses, one can compute and compare price curves for proportionate and progressive tax rates respectively, as shown in Fig. C.

If we use the following notations:

$V(t)$	= market value of house at time t	
a_0	= net annual user value (rent value) at time 0	
p	= rate of growth of user value	
s	= tax coefficient (tax paid in percentage of market value of house)	(40)
r	= discount rate	
n-t	= remaining economic life of house	
b	= parameter of tax progression, $s(t) = b\,V(t)$	

The market value of the house computed as the discounted value of future incomes and tax payments can then, with a constant proportionate tax coefficient, be written as:

$$V(t) = \int_t^n (a_0 e^{pu} - sV(u))e^{-r(u-t)}du \qquad (41)$$

which resolves into:

$$V(t) = \frac{a_0}{p-r-s} e^{pt}(e^{(p-r-s)(n-t)}-1) \qquad (42)$$

We now use the following parameter values:

$$\begin{aligned} a_0 &= 6 \\ p &= 0.08 \\ s &= 0.01 \\ t &= 0.06 \\ n &= 40 \end{aligned} \tag{43}$$

A computation of (42) with these parameter values gives the price curve I, in Fig. C. As expected the elasticity of price to changes in the tax coefficient is relatively low, -0,2, at the start and -0,1 at half-life.

Let us now introduce progression by setting s = 0.00007V(t). Compared to the current formal tax scales these rates are relatively low, both as to level and progression. Thus, they take some account of the effect of intermittent assessment. The market value of the house can now be written as:

$$V(t) = \int_t^n (a_0 e^{pu} - bV^2(u))\, e^{-r(u-t)} du \tag{44}$$

The explicit solution--which the common buyer is supposedly following in his evaluation--turns out to be a rather tortuous and long-winded expression.[1] The numerical result for the chosen parameters is shown as price curve III, in Fig. C.

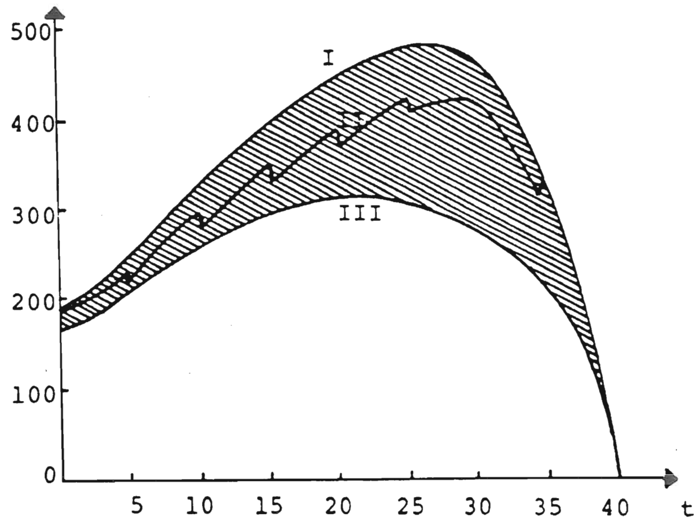
The elasticity of price to changes in the tax parameter is now very much higher, given a high rate of growth in user value.

---

[1] An account and discussion of the complete solution is given in Ysander (1976).

Figure C. Development of house price for different taxing and market behavior

House value
(Thousands Sw.kr.



Comparing the curves I and III we see that an increase of the tax yield is not likely to be the main effect of applying a progressive scale. First, and foremost, the price difference between the various categories of houses diminishes. Bigger and/or more comfortable houses become less profitable to build and sell.

Rather than taking full account of future progression, buyers and sellers may expect the current total tax coefficient to remain constant. The result would be a jumpy price development as demonstrated by price curve II in the figure. As shown by the Swedish experience in the seventies it is very difficult to predict when and how far tax rates will be adjusted for inflation or counteracted by assessing practices. The shaded area between curves I and III, in Fig. C, can be interpreted as a margin for the price uncertainty arising from progressive taxation. This margin will, moreover, tend to increase with inflation. The instability normally associated with changing inflation rates will thus be multiplied by this "tax uncertainty".

Taxation and wage inflation

Up to now we have dealt exclusively with isolated adjustments in a single market. However, the most widely observed and best known example of tax-induced instability relates to the adjustment of heavily taxed wage markets to price increases in the product markets, i.e., to inflation. This has been an acute problem in Sweden during most of the seventies.

In contrast to our previous examples we are faced, here, with annual tax adjustments aiming, mainly and explicitly, to compensate for the stability problems created by the tax structure itself.

The rates of income tax in Sweden are highly progressive -- and changing rapidly. Even excluding the various kinds of employers' social insurance fees, etc.--adding up to about 40 per cent of paid out wages--the marginal income tax rate for an average skilled industrial worker in Sweden now approaches 70 per cent, the average rate being some twenty per cent lower--all measured in terms of taxable personal income. The progression is steeper for high-income earners--and for low-income earners receiving subsidies.

If the worker, cited above, should be compensated for say a 10 per cent of inflation--with tax-scales not being automatically adjusted for inflation--he would have to receive a wage increase of some 17 per cent--starting off a run-away wage inflation spiral.

Negotiations are further aggravated by the variance in marginal tax rates between different groups of labor. Since gross wages are what is negotiated any compromise between the unions is likely to add further inflationary pressure.

Continuous tax revisions or an indexing of the tax scales provide the standard answer to the first problem--that of eliminating the "tax multipliers of inflation".

The second part of the problem however does not disappear so easily. Support for a tax redistribution of today's income does not automatically mean acquiescence in the further leveling of tomorrow's income implied by the marginal tax rates necessary to carry through the redistribution. To ward off this cause of wage inflation, annual revisions of relative total tax rates for various income-groups have, in recent years, become an important part of collective wage negotiations in Sweden. The structure of any progressive income tax is unfortunately such that every attempt to use tax revisions to satisfy claims for further leveling of net wages is apt to aggravate the "locking-in" effects and stability problems for the next round of wage negotiations.

There is another side of this instability problem that should be mentioned here, although it falls somewhat outside the model context of the previous discussion. Introducing progressive taxation, applied to gross market price, definitionally means, ceteris paribus, a lowering of the gross price elasticity of supply in the market. In terms of the labor market this means making labor less inclined to move in response to certain given wage inducements.

When this weakened pull effect is compounded, as in Sweden, with an institutionally and legislatively restricted push effect--by restrictions on how and when and why labor can be laid off--the possible consequences on market stability are apparent. The adjustment to shifts in foreign demand and/or to relative price changes will be slowed down and the competition for labor from expanding firms could either result in more inflationary wage increases or a petering out of expansion with inflated wage demand working as a damper.

Instead of Conclusions

Our previous discussion has involved a rather varied collection of examples of possible tax-induced instability. Our focus on individual market adjustment however means, that we have not treated the equally important problems of the impact of taxation on macro-economic stability.

The examples presented earlier do not readily lend themselves to any general interpretation or conclusion. They do however illustrate two important points.

The first one concerns policy. When raising the "technical" ambitions of tax-policy, gradually using it for more differentiated regulatory aims, the risk of disrupting the "normal" market adjustment processes grows.[1]

Stability problems are thus added to the more widely discussed problems of the long-term allocative effects of tax-induced changes in relative prices. The Swedish experience in the seventies seems to suggest that, also from the stability point of view, there are severe limitations to what you can safely hope to accomplish by tax policy.

The second point has to do with research. We have by now a fairly well-developed literature on "optimal taxation" and the welfare effects of a fixed tax structure from an "equilibrium point of view". Our examples demonstrate that there is now good reason to take one further step and investigate the impact of taxes and tax adjustment on market stability as well. Unfortunately, any thorough investigation into these problems will have to work with disequilibrium models, which makes points of departure harder to find. The results will also be less general and theoretically convincing. That may be an explanation for our being late to start but it is hardly an excuse for further delay.

---

[1] Alternative ways of pursuing these policy aims may of course be even worse from a stability point of view. The use of more direct intervention or regulation by definition makes the economy more rigid and hence less shock-proof. Having more "fixtures" and less free variability tends to narrow the margins of adjustment in the economy.

APPENDIX

THREE NOTES ON THE CONCEPT OF STABILITY

I Some basic stability concepts

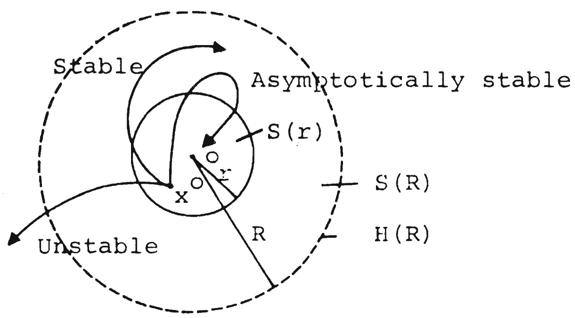
To facilitate reading the paper the reader may want to recall some basic stability concepts.

The concepts can be illustrated as in Fig. 1. We assume that we are dealing with an autonomous system, i.e., a system in which time, t, is not an essential variable but only used as a parametrization variable. We further assume that we are working in some open region of phase space, through each point, x, of which there goes a unique path of the differential system:

$\dot{x} = X(x),\ X(o) = 0;$

where x and $\dot{x}$ denote vectors.

Figure 1. Some basic stability concepts



We shall designate by S(r), S(R) the spherical region $\|x\|<r$ and $\|x\|<R$, respectively, and by H(R) the sphere $\|x\|=R$ itself.

We now say that the origin o is:

1) Stable (or stable in the Liapunov sense) whenever for each R there is an r < R such that a path initiated in S(r) always remains within S(R).

2) Asymptotically stable or resilient[1] whenever it is stable, and, in addition, every path starting inside some $S(R_o)$, $R_o>0$, tends to the origin as time increases indefinitely.

3) Unstable whenever for some R and r, no matter how small, there is always in S(r) a point x such that the path through x reaches the boundary H(R).

II Boundedness, Practical and Ultimate Stability

The usual basic concepts of stability analysis unfortunately turn out to be of little practical use when applied to price developments in real life economics. There are, in particular, four further problems that must be taken into account in any attempt at measuring stability in actual price movements.

In real economics time is an essential variable, i. e., the systems are non-autonomous. In theory, a generalization of the stability concepts to non-autonomous systems is straightforward although proofs tend to get more laborious. In practice we almost never know enough to analyze explicitly the time-dependence.

Resilience and stability are empirically indeterminate properties as long as we are talking in terms of some neighborhood which may be arbitrarily close to the origin. To acquire an empirical content the concepts must be quantified by measuring the extent of the regions involved in the stability definitions.

In most economic as well as physical systems, stability problems usually arise, not primarily because of initial conditions being far from equilibrium, but because of various kinds of persistent disturbances or perturbations. Any useful stability concept must therefore refer to the movements of such a perturbed system.

---

[1] Different authors use "resilience" to cover various shades or aspects of stability. We have chosen, here, to use the word when the system tends to become more narrowly confined within some neighborhood of an equilibrium.

Finally, we are often less interested in ascertaining the return to origin than we are in making sure that the system stays within bounds. Stability in the sense of Lagrange means just this, viz., that all solutions are bounded. Again this definition needs to be quantified to make empirical sense.

In trying to meet these four empirical requirements we could end up with the following two stability definitions that are illustrated by Fig. 2. Our starting-point is a system:

$$\dot{x} = X(x,t)+p(x,t), \quad t>0;\ X(0,t)=0 \text{ for all } t>0$$

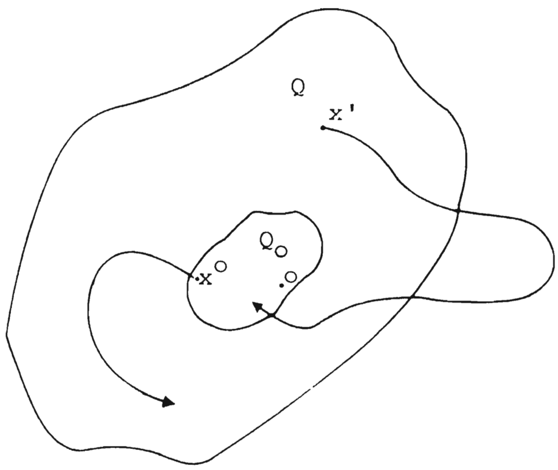
where p denotes perturbations satisfying $p<\delta$. We have, also, in the figure two sets: Q which is a closed and bounded set containing the origin, and Q_o which is a subset of Q. We could then, following LaSalle-Lefschetz (1961) define:

Practical stability of the origin as the property requiring that for given Q, Q_o and δ, any solution starting in Q_o will remain in Q for $T>t>0$ (cf. x' in Fig. 2).

Somewhat analogous to the concept of asymptotic stability or resilience would be:

Practical resilience: requiring that, for given Q, Q_o and δ, any path going through Q will be in Q_o for all $t > T$ (cf. x' in Fig. 2).

Figure 2. Practical stability and practical resilience



III. Structural and Comprehensive Stability

In most economic discussions of stability we deal with a system with fixed parameters where the path of prices, for example, can be completely described as a function of the state variables: $dx = f(x)dt$.

In real economies parameters do change. This is obviously the case with the parameters representing the state of the external world, such as world market prices for a national economy. Even if we simplify by ignoring these exogenously determined parameters we will still be faced with changing parameters.

In a widened or lengthened perspective we must take account of the fact that the behavior or the institutionally determined parameters of an economic system change according to some rule. Denoting the vector of parameters p, such a generalized explanation of change could be written as: $dx = f(x, p)dt$.

To avoid making the analysis too unwieldy economists usually try to discuss time developments in two stages - sometimes identified as a short and a long run. In the short run, parameters can be treated as given and the total change can thus be split into two parts:

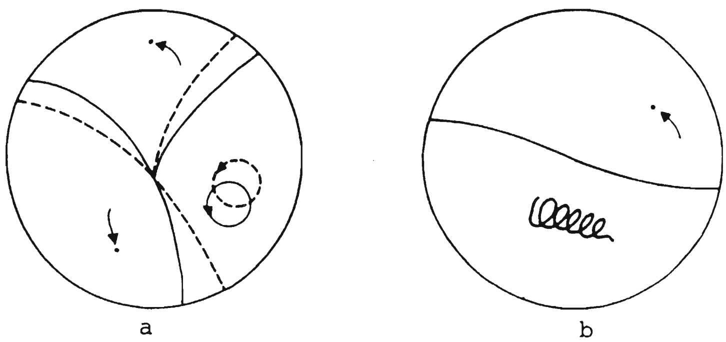
$$dx = f_1(x, p)dt + f_2(x,p)dp.$$

To be able to separate the impact of parameter change, f_2, from the "short-run" developments with given parameters, f_1, it is obviously necessary to assume that parameter changes are measured in time scales quite different from those used to define "short run" changes. This could be done by assuming parameter changes to be extremely "sudden". Usually however, economists go the opposite way, making the "comparative static" assumption that parameter changes occur slowly enough so that the "short run" system always has time to reach its asymptotic equilibria.

Instead of discussing stability as a property of the "phase-portrait", f_1, of a system with given parameters one may want to treat stability as a question of how big or how continuously the change in "phase-portrait" is, that results from certain parameter changes. This is roughly what is meant to be measured by "structural stability" in the sense of Smale (1967) or of the "catastrophy theory".

Fig. 3 may help to give some intuitive idea of this concept. Drawn with full lines in Fig.3a is the original "phase-portrait", which is supposed to be fairly simple--three basins, each with an attractor.

Figure 3. Change in "phase-portraits" caused by change in parameters



We now make a slight variation of the parameters and watch for results. The dotted lines in Fig. 3a show what could happen if the structure of the system is relatively stable. The parameter variation does not change the dynamic structure but only causes a continuous shifting of basins and limit-cycles. Fig. 3b illustrates a structurally unstable case where the same variation completely remodels the phase-portrait, reducing the number of basins and changing the character of attractors.

Once you include parameter changes in the framework of analysis there is one further question of stability to be considered. What causes parameters to change and does that kind of "system change" tend to counteract or reinforce instability "within the system"? Do institutions and economic behavior adapt in the long run so as to reduce or to maintain long-term imbalances? These questions concerning comprehensive stability --central to the current discussion of stagflation-- can, however, seldom be usefully analyzed within our economic models. The inability of our models to deal with "structural change" is indeed probably a major explanation for their poor showing during recent years.

REFERENCES

Arrow, K.J., Block, H.D. and Hurwietz, L., 1959, "On the Stability of Competitive Equilibrium", Econometrica, Vol. 27.

Cooper, R.N., 1967, National Economic Policy in an Integrated World, Council on Foreign Relations.

Foley, D., 1970, "Lindahl's Solution and the Core of an Economy with Public Goods", Econometrica, Vol. 38.

Grümm, H.R., 1975, Analysis and Computation of Equilibria and Regions of Stability, CP-75-8, IIASA.

---, 1976a, Definition of Resilience, RR-76-5, IIASA.

---, 1976b, "Resilience and related Mathematical Concepts" in IIASA Conference 10-13 May, 1976, Vol. 2.

Johansson, S.-E , 1978, "Hushållssparande och skatteregler", i Kapitalmarknadsutredningens slutbetänkande, SOU 1978:13, Bilaga 2-4.

Karlin, S., 1959, Mathematical Methods and Theory in Games, Programming and Economics, Vol. 1, Addison-Wesley.

Lancaster, K., 1968, Mathematical Economics, New York.

La Salle, J. and Lefschetz, S., 1961, Stability by Liapunov's Direct Method with Application, Academic Press, New York.

Lorentz, E.N., 1963, "The Mechanics of Vacillation", J. Atm. Sci., 20.

Milleron, J.-C., 1972, "Theory of Value with Public Goods: A Survey Article", Journal of Economic Theory, Vol. 5.

Mundell, R.A., 1962, "The Appropriate Use of Monetary and Fiscal Policy for International and External Stability", International Monetary Fund Staff Papers, Vol. 9.

Murata, Y., 1977, Mathematics for Stability and Optimization of Economic Systems.

Ruelle, D. and Takens, F., 1971, "On the Nature of Turbulence", Comm. Math. Phys., 20.

Smale, S., 1967, "Differential Systems and Structural Stability", Bulletin of the American Mathematical Society, 73.

Ysander, B.-C., 1976, "Skatteprogressiviteten och småhuspriserna", Ekonomisk Debatt, 4:5.

---, 1980, A Note on the Stability of Simultaneous Price and Adjustment, IUI, mimeo.

CHAPTER III

Homogeneity in Education. A Comment on the economic theories of education

Contents

Introduction

1 The one-dimensional man of economic theory

2 The homogeneity of human capital

3 Filtering homogeneous students: The Arrow model

4 Homogeneity in educational policy

5 From commodity-man to sequential machines

6 Education for variety

References

Introduction

During the first quarter century after W.W. II, most Western countries experienced rapid expansion of their educational systems, especially marked in higher education. This expansion was often supported by widely held optimistic assessments of the potential benefits of increased education in promoting economic growth and also in contributing towards an equalization of incomes. The human capital approach, developed during the same period by Chicago economists, provided a way of more strictly formulating the basis for these assessments and of incorporating the educational system into the main body of economic theory. Its conceptual framework made it possible - or so one hoped - to derive testable statements about the impact of education on productivity and growth and about the way the spoils of this growth were divided between individuals and between functional and educational categories.

In the last few years retarded economic growth in many countries has seemingly put a brake on both demand for higher education and the flow of tax-money available for educational expansion. At the same time the mood seems to have changed among both decision-makers and analysts. In discussions on educational policy, in West-european countries at least, one is now less apt to take the economic benefits of increased education for granted and more concerned with the qualitative direction of the educational system.

More than ever before there is an acute need for answers to the central questions about the economic role of education - what does education do to people, what do educated people do to production and what do they get out of it? So far economic analysis does not seem to have provided any clear-cut and wholly convincing answers in spite of the quite considerable empirical research work carried on during the last decades and the increasingly sophisticated refinements of human capital theories. The outcomes of attempts to measure the contribution of human capital to economic growth are still highly divergent and the methods very controversial. The distributional implications of the human capital theory have also been confronted with various kinds of empirical evidence especially in the form of estimated rates of return to educational investments. The results presented so far seem inconclusive and difficult to evaluate. While a considerable, sometimes major, part of income differentials can often be accounted for by educational differences, the existence in some countries of large divergences in the rates of return on different kinds of education and the unexpected degree of permanence in absolute and relative levels of these rates, even during periods of rapid educational expansion, seems harder to accommodate at least within the simpler versions of human capital theory.

There are also rival ways of interpreting the evidence. One such alternative interpretation that has lately attracted a good deal of attention views education mainly as a filtering device, whose main function is to measure and certify the given inherited ability of the students. That this interpretation cannot be just rejected as obviously conflicting with the given evidence is shown by the continuing laborious search - so far not very successful - for discriminating empirical tests to be used for choosing between the alternative theories. Since the policy-implications of the alternatives are in many ways radically different, this is certainly not a very reassuring state of affairs for the policy-makers.

At the same time developments in economic theory, especially the incorporation of uncertainty in general equilibrium theory and the outcome of the so called "capital controversy" in capital theory, seem to accentuate the need for a basic reassessment of the economic theories of education and a search for new approaches.

The aim of this paper is to make a modest contribution towards such a reassessment by focusing attention on one particular assumption in the current economic theories of education. This is the assumption of homogeneity - homogeneity of people and homogeneity of education - which we consider to be of strategic and pivotal importance in giving direction to educational analysis and research as well as to educational policy.

1 The one-dimensional man of economic theory

In general economic theory man appears in two very different roles. As consumer, property owner, manager and organizer the individual is treated as a very distinct personality set apart from his fellow beings by his own characteristic tastes and endowments of wealth and talent. Considered as a factor of production on the other hand he is usually reduced to an anonymous labor unit, whose hourly effort is indistinguishable from that of his fellow workers. The interest in economic theories of production and investment has traditionally been centered on real capital, which by being heterogeneous and immobile, supposedly defines the productive capacity of the firm or the economy. In contrast to this, labor - when defined as homogeneous, mobile and in given supply - is simply an amorphous commodity to be utilized within the chosen production technique. It regains an independent role only when attention is shifted from production theory to the theory of employment and capacity utilization.

This tradition of treating labor as a homogeneous factor of production goes back to the classical economists, for whom homogeneous labor units also provided a "measuring rod" for economic values. Abstracting from individual characteristics may also have involved less risks for misrepresenting the economic realities in the early industrial period when production in factories was organized simply, utilized mostly manual skills, and dealt with unorganized labor in abundant supply.

With the successive increase in the relative scope and complexity of the "control sphere" of production and in the amount of specialized education and on-the-job-training involved, the assumption of homogeneous labor becomes much more difficult to defend. As increasingly more heed is paid to the vested interests of different groups in the labor market and the aims of social security are widening to include security of work of a certain type in a certain place, the assumption of mobile labor also becomes more controversial.

The tendency among economists to fall back on an assumption of homogeneous labor is not just an historical accident [illegible] laziness. Instead it depends on certain inherent characteristics and limitations of the economic approach to social reality.

Economic theory is - at least in its central core - a theory about commodities and the exchange of commodities. Commodities can be anything in this world, goods or services, as long as they are completely specified, physically, temporally and spatially. Economic analysis usually presupposes a given well-defined commodity space. Every commodity - be it green figs or the services of an electrical engineer - is a given well-defined entity, the amount and price of which can be expressed in real numbers. From the point of view of economic theory each commodity is a black box with a given precise label. What is inside that box - the qualities and internal organization of the commodity - is not of concern to the economist, whose analytical toolbox does not equip him for that task. Instead that task is supposed to be handled by divers technical, or psychological, expertise.

Man as the object of exchange in the educational and labor markets makes up a special class of commodities - commodity man, who is supposed to have all the necessary educational and professional labels to make him a well-defined black box. Once he is so labelled we can supposedly treat him analytically as just another commodity, whose supply and demand and market price can be measured unambiguously.

Given the commodity space, the primary task of economic analysis is to find analytical expressions for the various and intercorrelated activities going on in the economy. For our interest here it is enough to point out three main categories of such activities, education, production and consumption. Here again the economist's ambition is limited. What really goes on within these activities, how the qualities of commodities there are created and appreciated, is something left to pedagogical, technological and behavorial specialists to analyze and express. The economist just takes it for granted that enough of this job has been done already so that he can start with a certain given description of education, production and consumption.

More specifically the description of technologies available for these activities should be given in terms of inputs and outputs of the well-defined commodities. This means the economist needs answers to questions of the following kind. If we put a certain number of men, with secondary school certificates and a given distribution of capability in some well-defined sense, through the existing process of higher education, what mix of doctors, engineers, economists and drop-outs can we get, how long will it take and what other commodities will be needed? How many engineers of different kinds do we need to keep a certain kind of paper process operation running at a given scale? How much paper of various kinds is needed for certain consumption activities, say small-boat navigation, and what is the output of "satisfaction" as measured in terms of some individual utility index? Given all these "technological" data in terms of commodities or functions of commodities the economist can get started on his special task which is the analysis of how all these millions of divers activities are intercorrelated and controlled, in the most usual case by way of market pricing.

This division of labor, the specialization of economic analysis, has almost certainly been a necessary condition for developing that unified body of analytical tools - general equilibrium theory - in which economists take a justified pride. It has indeed proved most effective, as long as economists stick to their primary task, the analysis of intercorrelated markets. The treatment of man as commodity-man, a well-labelled but otherwise unanalyzed entity, is then a well-motivated analytical conveniance.

Problems will arise,however, whenever the economist oversteps these self-chosen limits and tries to use his tools to analyze the technology and the technological change in education, production or consumption. To do so successfully he would somehow have to break into the black boxes with commodity labels to find out how they are organized , what qualities they possess and why they possess them. This is what technology is all about and technological change means among other things a reorganization of commodity components and a new mixture of physical characteristics. In the case of commodity man used as production input he would need to know e.g. what the elementary skills and capacities are that determine how good a certain professional is at his job and to what extent he can be substituted by some other kind of commodity man. If they wanted to analyze technological change they would also need to know e.g. in what way this vector of skill requirements depended on the machines used and on the general work organization in production, etc. To follow this up with a similar analysis of education they must find out how the training of these skills depends on various multi-facetted characteristics in the trainee and on the educational environment.

There are indeed economists that compete with other social scientists in studying these kinds of questions. But they cannot expect much help or guidance in this from the standard tools of economic analysis. Nor can they hope to arrive easily at results that are simple and general enough to be incorporated into standard economic theory.

What theoretical economists have normally done is to try and find a shortcut into technological analysis by way of simplifying assumptions. Since the paramount difficulty has to do with analyzing the heterogeneity of input and output commodities, the obvious way to try is to assume away most of this heterogeneity. We can e.g. try to describe production technology as if there exists only one kind of machines, one kind of labor, and one kind of output, although machines could be bigger or smaller and men more or less efficient. Even a truely staggering simplification like this seemingly gains some sort of credibility by a superficial association to the financial concepts of total labor costs and total capital costs. Should we then manage to estimate a production function in these simplified terms - and such estimates have indeed been made by the thousands although it is somewhat difficult to say exactly what has been estimated - we have also found a way to deal simply with technological change. We merely define this change as measured by the change in efficiency in either the machine or the labor unit or both. Technological change is said to be capital-augmenting, labor-augmenting or neutral. An outside critical observer of this whole procedure would probably be apt to think that the attempt rather shows the economist in the role of artful dodger. Technology and its change has been analyzed under the assumption that there is no technology in any real sense that can be changed. An assumption of homogeneity can of course serve equally well if we move from production to education technology. Since the output of education is the labor input to production, assumed to be homogeneous, there must be only one kind of educational results, although individual results, measurable in labor efficiency units or units of "human capital", can be larger or smaller. To arrive at a simple unified formula for education technology, without having to peep inside the black box of commodity-man, we also need a homogeneous input. The student then may be more or less smart, but his degree of smartness should be measurable on some linear scale, in terms of some given efficiency or capability unit.

The homogeneity assumption in education would appear to be at least as hard to support or interpret in terms of real life, as the analogous assumption in production, on which it depends. Not only does it abstract from all questions concerning the structure of educational technology. In as far as education is a way of sorting out people with divers talents and of making use of their comparative advantage for a labor market requiring a successively increasing specialization of skills any homogeneity assumption would appear extremely misplaced.

The one-dimensional man of economic theory thus becomes a very displaced person when he is designated as the central figure of a theory of education. An economic theory of education that attempts to analyze simultaneously the technology of education and production, to explain the role of education and of educated labor within production technology, without breaking into the black box of commodity man, would seem to be doomed to end up in meaningless abstractions or trivalities. To analyze the conditions and fluctuations in the markets for educated labor is one thing - and well within the scope of traditional economic analysis. To try to explain what happens to a man who gets educated, and how this affects his physical contribution in production is something else - and far beyond the limits of applicability for the standard tools of economic analysis. This is our contention here. Let us first take a closer look at the most widely used economic theories of education to see how they appear from this point of view. To what extent do they rest on an assumption of homogeneous labor?

2 The homogeneity of human capital x)

The dominant school of thought in the economics of education builds its explanations around the notion of human capital, as reinterpreted and revitalized in the pioneering works of Becker and Mincer; see Becker (1962, 1964) and Mincer (1962).

Research work based on human capital theories has, during the last decade, been extensive, branching off into many divers directions. This makes it by now somewhat difficult to examine the conceptual basis of "the" human capital theory, or, more specifically, to establish the exact meaning and pre-assumptions of the human capital concept. The author has elsewhere (Ysander (1977) , appended to this book of essays) reported on a modest attempt to explicate the notion of human capital in spite of these difficulties.

Both in economic literature and in common parlance "human capital" is used with many different connotations and in widely different contexts. One trivial but important distinction here is between the notion of physical and financial human capital respectively. Any individual, viewed as a potential source of some kind of labor services, can be said - in the economic jargon - to constitute a physical capital, since 'physical capital', definitionally, refers to something that yields services. That most economists still choose not to treat individuals as capital goods in this sense in their models is mainly a question of practical convenience. As long as you are focusing your interest on market operations and are not especially interested in the rather special category of educational investments, calling people capital goods in the model would simply mean renaming the individuals without getting any analytical gain.

People are not sold on markets or owned by firms like machines. Neither are e.g. the benefits of on-the-job-training offered on the market with a price tag. This means that if you want to treat individuals as capital goods and really study how they change by participating in educational and production activities your model must incorporate distinctions and transactions that have no explicit counterparts in real life markets. If you want to do this in a general way - meaning in a general equilibrium context as

---

x) In dealing with the problems in this section I have profited much from comments by Åsa Sohlman, who will present a more extensive analysis of the human capital concept in a forthcoming report within the same research project.

done in Ysander (1977) - it means i.a. assuming that households lease their human capital, i.e. their special labor capacity, to the firms under contracts which stipulate separate charges for the way these capacities are affected by the contracted work. In other words you must be able, always and everywhere, to account separately for all kinds of human investments and disinvestments. What you can gain by such a modelling effort is not really any new knowledge, but simply a consistent way of expressing yourself in discussing how an individual's working capacity develops over his life time. This general use of a concept of physical human capital certainly does not constitute any "human capital theory". It is more of a semantic convention.

The general financial concept of human capital is simpler and more straightforward. Any individual who can be expected to earn money in the future by selling his work services thereby constitutes a financial asset, whose value is the discounted sum of those expected future earnings. This financial concept of human capital can be used as a summary or shorthand notion in analyzing individual expectations and the way the individuals' choice of education and work may be determined by their earning expectations. This is again more a semantic convenience than a specific theory. "Human capital theory" could possibly be used as a kind of general name for all those theories about individual behavior in educational and labor markets that assume earning expectations to be an important determining factor. It would then be a very general name indeed. It can be shown that this financial concept can be worked into the general framework with physical human capital in such a way that you can define things like rate of return on human investments in a way consistent with the analogous definitions for non-human capital. This is not surprising and does not get us any closer toward a "human capital theory" in any more specific sense.

Any theory that is to have some substantial power of explanation must obviously be more than a way of framing concepts. Most of the research efforts of human capital theorists have been directed toward explaining various aspects of the income distribution in terms of human capital. What they are really trying to do is to reverse the reasoning that led to the financial concept of human capital. Instead of defining human capital as the discounted sum of future earnings, the task now is somehow to "explain" future earnings by means of a human capital concept that then cannot be financial but must be physical.

You are then really trying to explain the differential "productivity" of man. In terms of our previous discussion any such analysis and evaluation of the individual contribution to the joint effort in production would require opening the black box of commodity-man to determine the physical and mental characteristics that govern how well he or she operates in the technological organization. If you do not feel equipped for such a pioneering effort you can always resort to a homogeneity assumption, spiriting away the problems by way of definition.

There can be no doubt that this is essentially what is being done in modern versions of human capital theories, although the exact scope and content of the homogeneity assumptions may vary among the more sofisticated versions of the theory. x)

What assumptions have to be made in a simple and pure human capital theory, aimed at "explaining" relative earnings in terms of accumulated human capital - in terms of what first God and then men have invested in an individual - in such a way that the empirical information necessary for testing could in principle be distilled from existing market data? In what ways must a general equilibrium model with physical human capital be restricted to yield the desired type of relation? One possible way of answering these questions has been presented in Ysander (1977) and can be summarized in following manner.

First of all you should make sure that your explanation of relative earnings is really concerned with the physical productivity of the individual and is not just a way of rehashing the given data on market pricing. Relative earnings should depend only on the given amount of any existing forms of human capital, i.e. the given number of individuals with various earning capacities, but be independent of the rest of economy. This assumption can be shown to be equivalent to a necessary condition for aggregating human capital, for being able to substitute one number representing aggregated labor for the vector of differently skilled individuals in the description of the production technology. There is a certain irony in the fact that human capital theories, thus building on aggregating conditions, was first developed and circulated at the very same time, when - as an outcome of the

---

x) At the start of chapter 7 in Becker's book (1964) the author characterizes his own work in the following way: "Virtually all the implications of the theory of investment in human capital developed in Part One depend directly or indirectly on the effect of human capital on the earnings and productivity of persons and firms. Consequently most of my empirical work has been concentrated on measuring and assessing these effects."

That human capital must be interpreted as being physical becomes especially apparant when human capital theorists - - following the example set by Ben-Porath (1967) - introduce a "production function" for human capital (see e.g. the revised version of Becker's book).

The homogeneity assumption is also explicitly stated by Becker in his book (1964): "Another assumption made throughout most of the paper is that human capital is homogeneous in the sense that all units are perfect substitutes in production for each other and thus add the same amount of earnings."

so-called capital controversy - economists finally seemed to agree on the impossibility of capital aggregation in general.

Secondly, you must also assume that the human capital of different individuals is really all the same - just more or less of the same type of capacity. You can always use one man for another man's job but, depending on his relative amount of "human capital", you may then get a bit more or less done than before.

If "human capital" is to become something more than an empirically meaningless variable you also have to relate it to the various investments, training and job experiences, that have been made in order to form this capital. To avoid letting in heterogeneity by the back door you must assume some simple and common process for forming this homogeneous capital. If you want to keep it really simple you must assume that investments also are homogeneous - going to night school or gaining experience as a travelling sales man is really only more or less of the same thing.

Even after these assumptions you are still stuck with the fact that people are different and react differently to human investment efforts. If you want to reach an explanation of relative earnings that is quite generally applicable you somehow have to make people homogeneous. This can be done - and is usually done - by assuming that there is some unambiguously defined property called ability, with which individuals can be more or less generously endowed.

With these successive forms of homogeneity assumptions you can finally arrive at a general hypothesis concerning the relation between ability, investments and relative earnings. Whether you can test your human capital theory will thus mainly depend on the availability of valid and reliable data on this property called ability and on the various forms and measures for human investments.

Since there are many kinds of human capital theories there are certainly many relevant forms of homogeneity assumptions. What can be said generally is that in as far as human capital theories try to say something about the physical productivity of man they do so on the basis of far-reaching assumptions of homogeneity and can - in our view justifiably - be criticized on this account. They are trying to make summary conclusions about the role of man in production technology without really studying either man or technology; the homogeneity assumptions merely express the absurdity of any such attempt.

There is an alternative possible interpretation of the human capital theories. Perhaps "human capital" is not really meant to have any physical counterpart in reality but functions merely as an "intervening" or "theoretical" variable - a practical conveniance in giving a more general form to the empirical hypotheses concerning wages as partly determined by training investments? Becker and Mincer were after all colleagues of Milton Friedman at the University of Chicago. In a celebrated essay on economic methodology Friedman (1953) argued that the use of false or empirically meaningless postulates should be accepted as long as they lead to valid and interesting conclusions. (For an interpretation of this argument in terms of theoretical variables cf Ysander (1961)).

The Friedman argument has not found much support among economists and even less among the philosophy of science specialists. Even if it were accepted the burden of proof would anyhow lie with the proponents of human capital theories. They would have to show what general and testable conclusions are added by the introduction of a human capital concept. If the aim of the exercise is to show relative earnings as a well-behaved function of human investments or, more especially, of education, the use of a human capital concept does not seem to help to "explain" this relation in any real sense - but rather clouds the issue.

There are of course many alternative ways of explaining this relation. One such way might start with the idea that although people are different and function differently in production, employers seldom have any reliable means of controlling and measuring these factual differences; cf. e.g. Alchian-Demsetz (1972). By social convention they therefore take educational level etc. as one main starting-point in wage negotiations, since this at least is easily known and documented. They could then be said to practice a homogeneity assumption without having to believe in it. This hypothesis is just mentioned here as one of the many possible explanations that are open to you once you start interpreting human capital theories as theories about people behaving as if there existed a homogeneous human capital. They all have one common feature - that Occam's razor or the law of scientific parsimony would in most cases require you to avoid the use of homogeneous human capital as being both superfluous and cumbrous.

These objections to the use of a physical concept of homogeneous human capital should, however, not be construed as belittling the value of the empirical findings of research work carried out within human capital models. During the last decade this has helped us gain much useful knowledge about the correlations between distributions of human investments and distributions of associated earnings. What has been questioned here are not these data on the yield of various human investments but rather the attempt to interpret and "explain" these data in terms of homogeneous human capital.

3 Filtering homogeneous students: The Arrow model

The best known - or at least most talked about - alternatives to human capital theories in explaining educational impact on earnings are the so called filter theories of education, first presented in pioneering articles by Arrow (1973) and Stiglitz (1972, 1975). The central and rather provocative idea in these models is that instead of developing existing ability by investing the student with new skills, education merely functions as a way of certifying for the employers' benefit the given ability of the student. The rather depressing conclusions to be drawn from these premisses are that in as far as this certification merely affects distribution of income between employees without improving the allocation among jobs, education is simply a social waste.

Filter theories have usually been viewed as representing the extreme opposite to human capital theories. As we intend to show in the following, the two kinds of theories can equally well be seen as rather close - and from an empirical point of view often indistinguishable - substitutes, that both make the same basic assumptions. We use the Arrow model as the point of reference, as it is the simplest and most straightforward of the filter models presented so far. We present his premisses successively, interspersed with comparisons of their implications with those of human capital theory.

Arrow's filter model and human capital theory can be said to share two basic assumptions. First, people and educational processes are assumed to be homogeneous, although some students may be smarter than others. Secondly there is no generation of new knowledge through education, neither about the world around them nor about the students themselves. There is only a redistribution of already existing knowledge.

The homogeneity assumptions mean that as in the human capital theory, the filter theory totally abstracts from the technology of education and from the role of educated labor in production. While it attempts to explain why we have education at all, it cannot touch on the equally interesting questions of what determines the choice of a particular kind of education or a particular kind of job.

There is no search for new knowledge in the filter theory. Education is not a procedure for establishing the scholastic potentialities of the student, which are supposed to be closely, although stochastically, related to general ability or productivity. These individual probabilities of scholastic success are usually[x)] supposed to be known already to the student and, after admission scrutiny, also to the school authorities. All that education ever does is to certify for the employers something which is known all along by both students and schools. This means incidentally that the Arrow model cannot really explain why education is so time-consuming, why schools do not just pass on - for a price - the full extent of the information about the student they have gained on admission.

Compared to the human capital theories the filter model could be characterized by saying that its education system redistributes knowledge about the students to employers and the production sector, while in human capital theories the flow of redistributed knowledge goes the other way - knowledge about skills used in production are transferred from production to the students. The big difference is that, while in human capital theories knowledge of skills is assumed to make you more productive, the social value of information about students given to the employers in the filter theories are at best limited.

Does this in itself necessarily make the filter theories empirically very different from the human capital theories? Unfortunately this is not the case. Their conclusions about individual market behaviour could in fact be empirically impossible to identify separately.

What the alternative theories are both stating is that earnings - in terms of which a human capital can always be defined - are a function of ability and educational costs invested by the individual, (we abstract here from the fact that human capital functions have rarely been formulated stochastically as in the Arrow model). You could then in principle choose forms of investment functions and sets of investment possibilities for the individual such that the resulting pattern of investments and earnings in equilibrium would be the same in both kinds of theories. Then, from just studying an equilibrium solution in a real life economy, you could never tell the theories apart. For that you would need to dislodge the equilibrium, e.g. by rationing educational opportunities, and study the consequences on total production.

---

x) Arrow sometimes makes the alternative assumption that the students themselves do not know these probabilities (cf. e.g. p. 199 in the 1973 article). This, however, raises the problem - not discussed by Arrow - of how demand or self-selection for education is then determined and what can be known about demand as a sample of the total population.

One such way of bringing the theories closer together would be a kind of perfect filtering process, such that the longer you stayed in education, the higher the ability you could certify - assuming you had it. If wages were paid in proportion to certificates, and education costs were low, educational investments would then be monotonically increasing with ability, which is a common feature of many human capital theories etc.

Looked at from this point of view the provocation of filter models really arises from exploiting a general dilemma of economic equilibrium analysis. As long as we are limited to studying an equilibrium position of the economy in terms of market transactions, we usually cannot hope to distinguish between alternative dynamic explanations of how we have arrived at the equilibrium. Instead of framing the provocation in terms of humans you could equally well let it deal with, say, paper machines. You could make the hypothesis that paper machines are never really used for making paper, as naively supposed. They are instead simply a status symbol used to "certify" your capacity both to the world at large and to the marketing division of the company. You could undoubtedly go on to frame the hypothesis in such a way that it could never be refuted by just looking at an equilibrium solution. Only be getting far enough outside the equilibrium - or by getting permission to peep inside a paper plant - could you hope to settle this controversial question. In the case of the filter theories, you are certainly not supposed to be able to take a close look inside production to watch the performance of naked ability.

We have so far only dealt with those basic premisses in the filter model, that are so to speak generic to the filtering idea and thereby distinguish these theories generally from the human capital models.

The Arrow model has at least two other special traits. These are extremely important for his conclusions but do not necessarily separate his model from human capital theories, into which they could possibly be incorporated.

The first trait has to do with lack of discrimination on the part both of employers and educators.

Although educational authorities are supposed to know at the outset the individual scholastic indicators of relevance to productivity they are assumed to behave in a mean way. After having extracted a price from the students they only give the employers a small and rather distorted part of this information. What the employers get to know is only if the student has passed or not. For the successful students this means that the employers can make an estimate of the average scholastic potentialities, which is then supposed to govern their wage-setting for graduates.

One important consequence of this lack of discrimination, this averaging of certificates and wages, is that it can make education profitable also for some less able students who, as graduates, will be paid above what their ability would justify in a more discriminating world. It thus introduces a new source of possible inoptimal allocation into the model.

This assumption about a lack of discrimination in the labor market could in principle also have been superimposed on a human capital model, although it would undoubtedly detract from the formal elegance and simplicity of the theory. It is an assumption about the state of information on the market, which could be combined with many alternative explanations of productivity.

To illustrate that filters may also have a socially beneficient function, Arrow also uses a second set of assumptions about the labor market. Filtering students will obviously improve allocation - if we abstract from educational costs - when the results can be used to assign students between different jobs in such a way that their ability is better utilized. In a model with homogeneous students and homogeneous education it is rather difficult to introduce heterogeneous jobs in any real sense. Arrow avoids this difficulty by assuming a very special segmentation of the labor market. Half of the total utilized productivity of labor must be used for jobs in which only a minimum of potential productivity, common to all labor, can be utilized by each employee. We can visualize these jobs e.g. as manual tasks of operating simple machines, such that any individual can manage one machine but none can cope with more than one for physical reasons. Filtering by education can then help in assigning less able people to these simple but tedious tasks. What is assumed is thus a special combination of indivisibility and complementarity in the labor market.

What interests us here is not whether the assumptions can be said to model any relevant features of real life - which may be doubtful. The point to be made is instead that these assumptions are not in any way necessarily related to the filter hypothesis, although they determine the possible social benefit of an educational filter. Similar assumptions - formulated as limited possible yield of human capital in certain employments - could evidently also be introduced into human capital models.

Arrow's conclusions and evaluations about the equilibrium amount and distribution of educational investments all depend on his special "labor market assumptions". Other assumptions would of course lead to other conclusions. We could e.g. assume that the labor market were partitioned into as many segments as ability, with each segment only allowing the use of a certain limited ability, and that education could function as what we above called a perfect filter, successively filtering out higher degrees of ability. If the marginal productivity gain and wage increase from more education were everywhere larger than the marginal educational costs there would be no risks of inoptimality in equilibrium.

The results of our discussion could perhaps be summarized in the following way. The filter and human capital approaches are usually seen as dramatic contrasts. They could equally well be described as rather close substitutes, which share many decisive basic assumptions, including that of homogeneity. The provocative difference in conclusions - if we abstract from other superimposed assumptions about the labor market etc. - arises from different interpretations of the productivity of educated labor, in a given equilibrium situation. What makes this difference especially provocative is the fact that it may be empirically impossible, within equilibrium analysis, to decide which interpretation is right.

A much more radical departure from current orthodoxy would be a model of search by learning, which incorporated the heterogeneity of students, educations and jobs and the generation of new knowledge in education, by analyzing the educational process as a search for the genuinly unknown qualities of the students. If the only way to find out what you are good at is by learning to do different things, with various degree of success, then the conflict between the alternative interpretations of productivity would also tend to disappear. If ability, as measured on the scales of intelligence tests, merely gives the length of your capacity vector, without telling the direction - which could be even more decisive for your productivity in a special job - then there may not be any competition between the claims of ability and the claims of training. If education is partly a way of "getting to know yourself" as the ancient Greeks believed, then the form and structure of the educational search is all-important and the educational choices for the individual much more complex than what is modelled in current theories.

We mention this possibility here only to emphasize our conjecture, that homogeneity versus heterogeneity is a much more strategic choice for the direction of any future economic research in education than the human capital versus filter interpretation.

4 Homogeneity in educational policy

The homogeneity assumptions in the economic theories of education may not only distort the theoretical conclusions but also lead to misunderstandings in educational discussions and to mistakes in educational policy.

There exists an unfortunate tradition of mutual misunderstanding and disrespect between economists and sociologists. Real communication and cooperation between them nowhere seems so hard to achieve as in the field of education. (There are many outstanding exceptions to this rule - one being our co-author in this volume - Mary Jean Bowman). In view of our earlier discussion of the homogeneity assumption this is hardly surprising. Sociologists, who aim at establishing the differences in people - in background, experience and mental characteristics - that determine their choices of different educations and jobs, must surely find it hard to pursue a meaningful discussion with economists who start by assuming away all, or almost all, relevant differences. In the same manner there seems to be an obvious lack of common ground for economists and pedagogical research workers as long as economists insist on treating as both homogeneous and irrelevant the black box of the human mind, that pedagogical research aims at analyzing and manipulating.

Reasoning about education on the basis of homogeneity assumptions, however, is not a special prerogative of economists. Many policy decisions in the educational field - we refer especially to Swedish policy for higher education in recent years - seem to be based on reasonings of a similar kind.

Discussions about the organization of higher education are often focused on two rather different models, central rationing and decentralized marketing of educational opportunities. Homogeneity assumptions appear to be pivotal in weighing the decision in favor of central rationing.

If you are willing to act as if the homogeneity assumptions were true, i.e. as if students were all the same - only some smarter than the others - and indifferent between various kinds of education this means that students, teachers and local school authorities have no information relevant for distributing various educational opportunities that is not also easily available to central authorities. These authorities, on the other hand, should be able to interpret signals about labor market demands, both in the short and in the long run, better than any one individual in this standard collection of students. Social efficiency reasons then weigh in favor of central rationing. So as a matter of fact do equity reasons as long as equity can be identified with equal educational and labor market status.

If instead you start from the contrary assumption, that students are fundamentally different in kind and in their aptitude for different sorts of educations and jobs, you will be faced with information problems of quite another dimension. There is so much more you now need to know about each individual in order to channel him or her into the right kind of education and the right kind of job - and much of this information may be available even to the student himself only after a laborious search process. For the same reasons the segments and aspects of labor market information which are relevant will now differ between the individual students. Instead of having almost a corner in relevant information, central authorities now appear to have lost all comparative advantage in distributing opportunities and in steering the individual student through the maze of courses and crafts. With this starting-point you will then tend to favor a decentralized marketing of education services with quality competition for students between the various educational organizations, leaving it to the student himself to interpret and react on market signals both from education and from the labor market.

Even the equity goals may look different from these premisses. If students are really that much different it does not make sense to define equity simply as having an education or a job of equal status with the rest. Strategically important aspects will now be how apt and motivated you were for the kind of education you got, which in turn will determine how well you do and how adjusted you will be in the job your education prepared you for. Equality in education must then be treated separately and independently of equality in the labor market. Equal opportunity to search for and find what you really want to do and feel good at - should constitute the equality aims of education. It will lead to social equality in a more comprehensive sense only if combined with the equalizing of status and pay in the labor market which however requires other kinds of policy instruments. To try to equalize the labor market by way of educational policy is anyhow not only ineffective - at least in the short run - but would also appear to be a rather half-way kind of ambition, since it means that you want to adjust the educational system to the traditional unjustices in the social evaluation of different jobs.

Whether the prevalence of homogeneity assumptions in policy discussions is to some extent due to the impact of economic reasoning or if it just happens that such ideas come naturally and spontaneously to the bureacratic mind, is difficult to know. What our examples above show - if somewhat obiquely - is anyhow that economic theories of education, if taken seriously, could have important - and in our view disastrous - implications for educational policy.

5 From commodity-man to sequential machines

We have so far only discussed the two dominant theoretical themes among the economic theories of education, both of which were found to be based on homogeneity assumptions. This does not mean that there has been no attempt by economists to break into the black box of commodity-man. Such attempts have been made in various directions, although so far without any definitive breakthrough either in concepts or in empirical measurements.

The importance of differentiating, in terms of quality the analysis both of students and education beyond the one-dimensional ability variable has been stressed by many writers who have dealt with education in relation to the labor market; cf. e.g. Blaug (1966) and Rees (1971). A few economists have even tried, theoretically, to define the individual explicitly in terms of a vector or profile of qualifications, which are changed by education and which determine his usefulness in different employments. By so doing you not only gain a way of discussing the students' comparative advantages, when faced with heterogeneous opportunities of education and work. You also overcome a traditional handicap of economic analysis by being able to "explain" and not only register relations of substitution between students and between jobs and the effects of introducing new kinds of education or new kinds of jobs.

Tinbergen (1963) and Mandelbrot (1962) are perhaps still the two best known examples of economists, who have tried to use this kind of model to explain relative earnings. Tinbergen assumes demand for employees in different jobs to be specified in terms of required profiles and then derives a wage function, common to the whole labor market, from the common utility function of the individuals. Mandelbrot goes the opposite way, assumes a separate wage function as well as a totally elastic demand of labor within each job category and then studies the result of simple wage-maximizing by the individuals, who are supposed to be distributed among different profiles in a well-known way. Neither deals explicitly with education. A natural way of analyzing education in these kinds of models would be to describe it as a way of simultaneously changing and gaining knowledge about the originally given profile of an individual.

Further progress in this direction seems so far, however, to have been blocked by the lack of empirical data for interpreting the "profiles" in the models. Early hopes of being able to use e.g. military service records and the results of currently made so called requirement analyses for different jobs, have all been frustrated. These records turn out, in most cases, not to measure - at least not in a systematic fashion - the kinds of functional qualifications and attitudinal properties that would be needed. This unfortunately leaves us with elegant theories, whose concepts remain empty.

Instead of hoping that knowledge of what are the relevant dimensions will somehow be furnished from someone else - work psychologists perhaps - some economists have recently attempted at least to structure the problem by analyzing the intellectual functions involved in controlling production.

One simple intuitive idea behind these attempts is to draw advantage from the fact that we are living in the computer age. If man can be at least partly replaced in an increasing number of jobs by sophisticated computors, then, surely, we should be able also to do the reverse and analyze some of the main intellectual functions of man in his role as organization man by studying the network of mechanical components that could replace him. Even a rather simple clerical task to be replaced, requires a sequence of elementary computor units with capacity respectively to, say, receive signals, decode and interprete, memorize, apply decision rules, calculate and evaluate consequenses, code and transmit signals, etc. By studying man as sequential machines in different job situations, we might hope to pinpoint some of the main dimensions of intellectual capacity, important for the individual contribution to work in an organization. Machines, however, can be self-organizing only to a certain limited extent and therefore we cannot ever hope to catch all the more creative facets of the work effort. It may anyhow provide a starting-point for what really interests us here, the analysis of how these capacities can be acquired by variously gifted people through education.

Starting from this idea, it seems natural to apply the same method also to the analysis of a whole organization. A pioneering attempt in this direction has been made by Radner and Marshall in their well-known work: Economic Theory of Teams" (1972).

The theory of teams studies organizations as networks of component units for receiving and interpreting information, for computation, for the application of decision rules and for the execution or the transmitting of orders. It can be used to calculate e.g. optimal decision rules and/or optimal information structures for a given organizational network, whose aim or pay-off-function is known. From these calculations you can go on to compare the efficiency of alternative networks, when optimally utilized.

These rather abstract notions can be given some intuitive content by way of a simple and extremely stylized example.

Let us think of a shipping company that sometimes finds it profitable to operate in two special freight markets - here called markets 1 and 2 - where freight rates tend to diverge markedly, both in positive and negative directions, from rate levels in the company's accustomed market, where they are supposed to be able to forecast developments accurately. If they engage in market 1 - with a fleet of a certain given size - and relative rate developments are favorable - let us represent this as $x_1 = 1$ - they stand to make an added profit of α millions of dollars, while unfavorable developments, $x_1 = -1$, will result in a symmetrical relative loss of the same amount. They have only two alternatives in this respect; either they decide to engage - represented as $a_1 = 1$ - or they do not, $a_1 = 0$. Their situation when it comes to market 2 is completely analogous, although here the potential profit or loss amounts to β. There is one complication, however. If they try to engage in both markets simultaneously they will run out of ships and will sustain an extra cost of c millions for hiring the required extra tonnage. We assume that $\alpha > c > \beta$ which means that it can never be profitable for the company to hire outside tonnage.

These assumptions can be summarized in the following pay-off-function which simply states the total relative profit, resulting from the company's actions in these markets:

$$\Pi = \alpha x_1 a_1 + \beta x_2 a_2 - c a_1 a_2$$

Several types of management functions are required to handle this problem. An observation function, 0, is needed, to "read" or forecast the relative rate development in the respective markets. On the basis of these forecasts, a decision function, D, must make decisions on whether to engage or not. Finally these decisions must be realized by an executive function, E.

Three alternative networks for combining these functions are depicted in Fig. 1.

The first alternative, Fig. 1a, means that both forecasts and decisions are made centrally and without using specialist knowledge on the respective markets. The price for this lack of specialized knowledge is represented here by an error in forecasting, ε, which with probability q, takes on the value, -1, i.e. makes both market forecasts misleading, but is otherwise equal to 1.

In the second alternative, Fig. 1b, two specialist observers are used, supposedly making the risk of error negligible. Their reports, however, are still fed into a central decision unit etc.

The third alternative differs from the second in that decisions are also made on a decentralized basis by the specialist observers, which introduces the risk of incurring extra costs by having to hire tonnage.

With specified assumptions on the probability distributions involved - for x_1, x_2 apart from ε - definite conclusions can be derived about the comparative advantages of these alternative networks, when utilized in an optimal fashion.x)

What is of special interest to us here is the possibility of deriving organizational demand for various types of intellectual capacities as a function of optimal organizational structure, which in turn will be determined by organizational aims and costs and by environmental conditions.

In our simple example we can study, for instance, the organizational demand for specialist freight market observers mainly as a function of the stochastic properties of the special freight markets involved and of the organizational costs of centralized decision-making. We can thus estimate the value of the contribution or "marginal productivity" of these specialized capacities under varying environmental and organizational conditions. We have then taken a first step towards "explaining" how certain acquired intellectual capacities contribute to the joint output or organizational pay-off-why learning may motivate earning.

---

x) If we assume that the two alternative values of x_1 and x_2, 1 and -1, are equiprobable with correlation coefficient r, elementary calculations show that the expected gross value of alternative c is:

$$E\Pi_c = \frac{1}{2}(\alpha + \beta) - \frac{1 + r}{4} c$$

A change over to alternative b - centralizing decisions - increases relative gross profits by the following positive amount:

$$E\Pi_b - E\Pi_c = \frac{1 + r}{4}(c - \beta)$$

The effects of a further change into alternative a would decrease the expected relative gross profit:

$$E\Pi_a - E\Pi_b = -(1-q)\left(\alpha + \frac{1 - r}{2}\beta\right)$$

These gross profit figures must then be combined with the organizational costs for the various alternatives to arrive at conclusions about the most profitable network.

Figure 1a. Centralized observation and decision

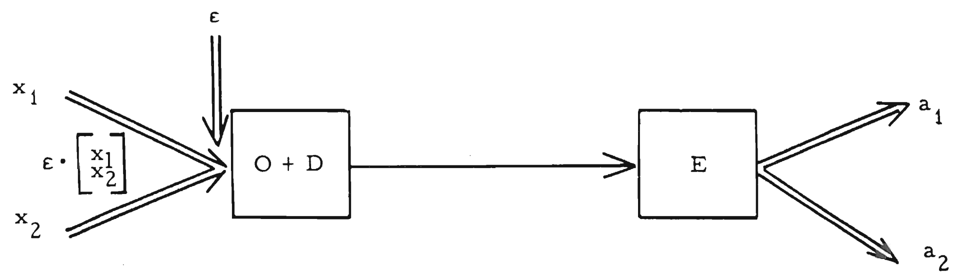


Figure 1b. Decentralized observation and centralized decision

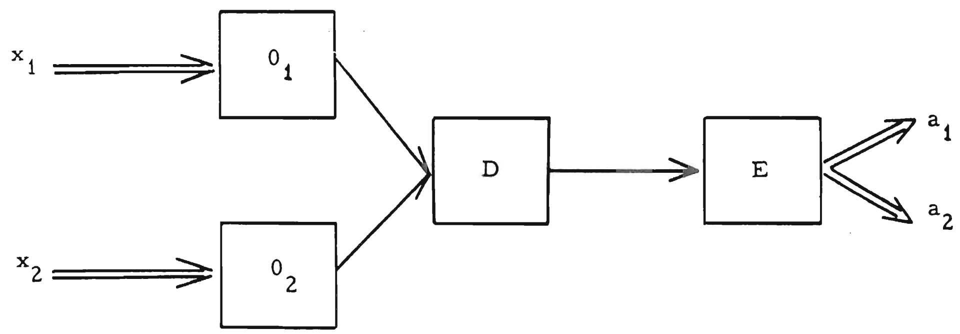
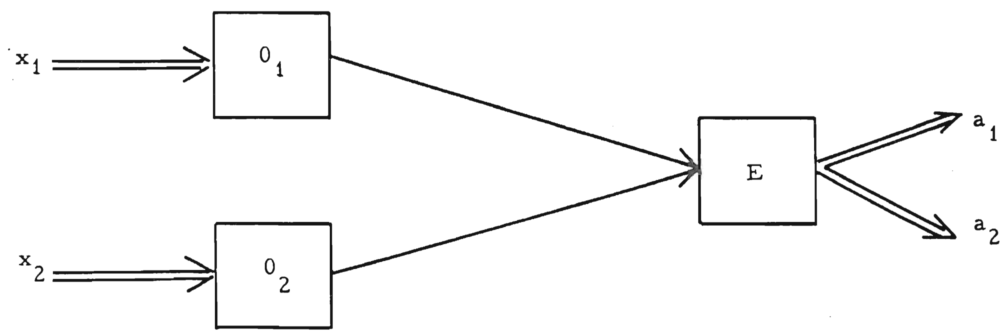


Figure 1c. Decentralized observation and decision



What one would ideally like to envisage is an interpretation of these "computer-capacities" as psychologically measurable categories, so that a start could be made in really analyzing the ways in which education makes people "more productive".

It should be admitted, however, that so far we have little basis for any great hopes in this respect. The theory of teams and related approaches are still only abstract conceptual schemes, whose empirical usefulness still has to be proved. Moreover the capacity categories used are still probably much too general to make a psychological interpretation possible. But the approach does represent a rather unique attempt to break into the black box of commodity-man in production.

Even if this attempt should prove successful we have only gone part of the way towards an economic theory of the role of education in production. Man is undoubtedly more than a sequence of machines, he is also a living organism, a complex of sometimes conflicting motivations as well as a bundle of creative instincts. The effects of education on productivity must probably be analyzed also in these terms since most changes in the educational system tend to change the psychological and social conditions for study.

6 Education for variety

The conclusions of our discussion of economic theories of education have been mainly negative. It was suggested above that economists have been forced into making homogeneity assumptions when trying to use their analytical tools for problems outside their traditional and legitimate field of market studies - in trying to probe into the "interior" of educational and production processes. By using this "short-cut" their results have also in our view been rendered rather useless but unfortunately tend to support the corresponding policy assumptions which may have far-reaching effects on the shaping of educational organizations. Attempts to develop new analytical methods for the analysis of the "productivity" of educated labor have been made but have not yet been developed far enough to hold definitive promises. Our own conclusion from this would be that economists still have to be very modest in their claims of "explaining" the effects of education in production. They can claim to have real expert knowledge only as long as they stick to those relations in the labor market which can be verified from market data.

We know that the homogeneity assumptions, taken in a literal sense, are false. To substantiate our criticism above we would however need to know how wrong they are in relevant respects; something which unfortunately we cannot know in the present state of research. For the present we have to fall back on subjective beliefs and attitudes. If we start with a one-dimensional view of our fellow-men, it will be consistent to view education as a way of spoonfeeding the test-tube babies in "1984" and educational policy as mainly a question of choosing the spoons. If we base our beliefs on an explicitly pluralistic conception of man-kind, we arrive instead at a conception of education as a way of finding and developing the special talents and motives of each individual within the restrictions given by production technology. The aim will then be an education for variety which also means a variety of educations and a corresponding variety of signals and signposts to make it possible for the individual to make rational choices in each successive step of learning and earning.

References

Alchian, A.A. - Demsetz, H. (1972): "Production, Information Costs and Economic Organization", American Economic Review, vol. 62, no 5

Arrow, K.J. (1974): "Higher education as a filter", Journal of Public Economics, vol. 2, no 3

Arrow, K.J. (1974): The Limits of Organization, N.Y.

Becker, G.S. (1962): "Investment in Human Capital: A Theoretical Analysis", Journal of Political Economy, Supplement, Oct.

Becker, G.S. (1964): Human Capital, N.Y.

Becker, G.S. (1967): "Human Capital and the Personal Distribution of Income": W.S. Woytinsky Lecture, No 1

Ben-Porath, Y. (1967): "The Production of Human Capital and the Life Cycle of Earnings", Journal of Political Economy, vol. 75, No 4

Friedman, M. (1953): Essays in Positive Economics, University of Chicago Press

Mandelbrot, B. (1960): "The Pareto-Levy Law and the Distribution of Income", International Economic Review, no 1

Mandelbrot, B, (1960): "Paretian Distributions and Income Maximization, Quarterly Journal of Economics, vol. 76

Marschak, J. - Radner, R. (1976): Economic Theory of Teams, Yale University

Mincer, J. (1962): "On-the-job Training: Costs, Returns and some Implications", Journal of Political Economy, Supplement, Oct.

Mincer, J. (1970): "The Distribution of Labor Income, A Survey with special Reference to the Human Capital Approach", Journal of Economic Literature, vol. 8, No 1

Radner, R, (1972): "Teams", in Decision and Organization, (ed. McGuire, C.B. - Radner, R.), North Holland

Radner, R. (1972): "Allocation of a scarce resource under uncertainty: an example of a team", in the same book

Spence, M. (1976): Market Signalling, mimo, Harward University

Stiglitz, J.E. (1972): "Education as a Screening Device and the Distribution of Income", mimeo, Yale University

Stiglitz, J.E. (1972): "The Theory of "Screening", Education and the Distribution of Income", American Economic Review, vol. 65, no 3

Tinbergen, J. (1956): "On the Theory of Income Distribution", Weltwirtschaftliches Archiv, Bd 77

Tinbergen, J. (1975): Income Distribution. Analysis and Policies, Amsterdam

Ysander, B.-C. (1961): "Truth and Meaning of Economic Postulates' in Money, Growth, and Methodology, (Ed. H. Hegeland), Lund

Ysander, B.-C. (1977): "The Meaning of Human Capital", in this volume.

Appendix

Bengt-Christer Ysander

THE MEANING OF HUMAN CAPITAL

Contents

Introduction

1 Human capital and general equilibrium
1.1 The context: A myopic general equilibrium
1.2 Production and consumption
1.3 The holding companies
1.4 Factors of production
1.5 Completing the specification
1.6 Income, wealth and the rate of return on human investments

2 Homogeneity of human capital
2.1 The need for further assumptions
2.2 The first assumption: Physical human capital and the possibility of aggregation
2.3 The second assumption: Homogeneous human capital
2.4 The third and fourth assumptions: Homogeneity of human investments and the ordering of human types

A note on stationary equilibrium

References

This paper reports on some work carried out within the research project: "Education and the Labour Market", sponsored by the Chancellor's Office at the University of Stockholm. The problem treated here was originally taken up by one of my collaborators on this project, Åsa Sohlman, who intends to present a more extensive analysis of the human capital concept in a forthcoming report. I have profited much from her comments and suggestions.

Introduction

The idea of human capital is as old as the subject of political economy itself and the writings of classical economists are often interspersed with references to the economic analogy between men and machines. In the last two decades, following the seminal writings of Becker, much interest and research effort has been centered on the possibility of building empirical hypotheses around the notion of human capital, aiming e.g. at explaining relative earnings in terms of educational investments.

It is hardly surprising that in this sudden outpouring of writings and studies, many diverse and mutually inconsistent notions of human capital have been exemplified. "Human capital" has usually been treated as a primitive concept, whose exact meaning and relation to other concepts of economic theory is often left unspecified.

Human capital theory has by now matured to the stage where there is a plethora of empirical results to be evaluated, a rising number of competing alternative theoretical approaches and consequently a growing need for a critical appraisal of the central theoretical foundations of the theory.

What is the exact meaning of "human capital" and how should it be fitted into the general equilibrium theory on which human capital theory is supposed to be based?

This paper aims at giving at least a partial answer or one possible answer to these broad questions.

The discussion in the following is divided into two chapters.

The aim in the first chapter is simply to try to fit a very general concept of human capital into the framework of equilibrium analysis, without heeding the more specific needs and notions of so-called human capital theory.

If we really want to treat men as machines, i.e. as heterogeneous capital goods, taking part in production and being changed by production in various ways, how can this be formally represented? In contrast to machines, men are not tradable and investment in men means investment not in production but in change.

It will be shown that to make room for separate accounting of human capital and human investments the usual equilibrium framework must be modified both as to institutional assumptions and the specification of commodity space.

Once this is accomplished human resources can be socially accounted for in analogy with machines or buildings. This means, as it is spelled out in section 1.6, that we can have a human capital theory in the sense of a financial account for human capital and investments following the usual accounting rules for profit rate and rate of return on capital in equilibrium.

This may be enough to explicate the meaning of classical references to human capital in general but it certainly does not provide a basis for human capital theory in the modern sense of Becker. The second chapter is dedicated to an attempt to give a stepwise account of the further restrictions on the general equilibrium model, that are needed to provide a notion of human capital useful for the purpose of explaining relative earnings.

Searching for conceptual clarification is seldom exciting and often rather tedious. It is, however, a necessary starting-point for any critical reappraisal of a theory's viability and validity.

1 Human capital and general equilibrium

1.1 The context: A myopic general equilibrium

The Arrow-Debreu model of general equilibrium as presented in Debreu (1951), Arrow-Debreu (1954) and Debreu (1959), has a well-known intertemporal interpretation involving the assumption of dated commodities and of a complete set of spot and forward markets.

The general equilibrium framework we are going to use here is a respecified, reinterpreted and somewhat extended version of this intertemporal Arrow-Debreu model.

The production possibilities of the firms will be presented disaggregately and specified period by period under the assumption that outputs of a particular date depend only on inputs at the preceding date. This way of representing production by period in terms of a generalized activity analysis was first introduced by Malinvaud (1953 and 1972) and has later been used, discussed and developed in Bliss (1975).

The intertemporal equilibrium will also be reinterpreted as a temporary equilibrium, existing in each consecutive period. In the Arrow-Debreu model the economic agents make transaction decisions once and for all in the first period, after which they are faced with an over-all budget restriction and have access to a full complement of forward markets. In a temporary equilibrium model the agents are assumed to make decisions concerning transactions only in the current period. They only have access to spot markets including a bond market linking current transactions with the future. The relevant budget restriction also holds for current transactions only. From current prices and interest and expectations about future prices and interest the agents decide on an optimal allocation of their budgets. A temporary equilibrium will ensue when these decisions are such that supply equals demand on spot markets and the bond market.[x)]

---

x) Early examples of the use of a temporary equilibrium concept are provided by Lindahl's classical paper from 1929 and Hick's discussion of a "Spot economy" in his 1939 book: Value and Capital. A survey of temporary equilibrium models and of the contemporaneous efforts to extend these models to deal with situations involving uncertainty and quantitative restrictions is given in Grandmont (1976).

If we assume that agents have perfect foresight and that their plans are fully consistent so that their price and interest expectations will all come true, we get a very special kind of temporary equilibrium or rather a sequence of such equilibria. This is very close to the original Arrow-Debreu model since in both cases the agents make all decisions with full knowledge of the true development

For our purposes here the use of a model of temporary equilibrium with perfect foresight affords the advantage that definitions and discussions of human capital concepts can be made simpler and more intuitive by being framed in terms of current period decisions. By suitable assumptions on the firm's financial arrangements the firms will also be seen to behave myopically in this model, i.e. they will make their decisions on the basis of current input-output possibilities only and will endeavor to maximize current dividends. That is the reason why we have chosen to call the model myopic general equilibrium - in the following abbreviated as MGE.

But if we wish to pursue the discussion in terms of an arbitrary period we must also remove the asymmetrical treatment in the Arrow-Debreu model of the first and last periods. This will be done - following Svensson (1976) - by allowing for initial debts in the first period and by letting the model extend into an infinity of future periods. There is, however, a price to be paid for this pedagogical convenience. Although equilibria with bankruptcy and with an infinity of commodities have been examined recently there can be no assurance here of either the existence or efficiency of this extended version of the model. A further extension in comparisons with most standard equilibrium models will be made here by allowing for the possibility of negative prices for some special goods. How this can be done and what it implies has already been shown - first by Arrow (1951) and Koopmans (1951).

As an obvious safeguard against perverse cases made possible by the infinite horizon we also have to assume convergence for the various discounted values involved. This will in particular exclude the case where expenditure is financed by indefinitely postponing the payment of an ever-increasing debt.[x)]

---

cont.

of prices and interest over time. It differs solely in the respect that the over-all budget restriction in the Arrow-Debreu model is now broken down into a series of temporary restrictions linked by the bond market, forcing the agents to postpone their transaction decisions in a corresponding manner. It can also easily be shown that the perfect foresight temporary equilibrium is equivalent to the Arrow-Debreu model in the sense that any consumption and production allocation in one model can also be realized in the other. This has been shown by Guesnerie - Jaffrey (1974) for the exchange economy and by Svensson (1976) for the economy with production.

---

x) The MGE-model here used is very closely related to Malinvaud's model (1953) and its specification has borrowed many traits from Svensson's paper (1976). It differs from Svensson's model mainly in the special institutional arrangements and the more general treatment of non-produced factors of production used here as a means of explicitly introducing human capital.

1.2 Production and consumption

In presenting the model we begin with its more general features; the special characteristics needed to deal with human capital will be introduced in the next section.

The economy has I consuming households and J production firms. In addition to this each household is assumed to own and operate one of I holding companies. The raison d'être for these holding companies will be given in the next section.

There are at each date at most U goods, including both commodities and services, both produced goods and non-producible factors. Time extends over $t = 0, 1, 2 \ldots$ up to infinity. We want to study the economy at an arbitrary date t. For the later specification of the full intertemporal equilibrium, however, it is most convenient if we choose to study the economy at $t = 1$.

At any chosen time $(t \geq 1)$ there are spot markets for all of the U goods and also a bond market, with prices, $p_t \in R^U$, and interest r_t.

A household i at time 1 can be characterized by a quadruple $(C_i, \precsim_i, \nu_{io}, c_{ibo})$ where $C_i \subset R_+^{\infty}$ (R^{∞} is used here as an abbreviated notation for the countably infinite sequence $R^U \times R^U \ldots$) is the consumption possibility set of which the elements

$c_i = (c_{i1}, c_{i2} \ldots) \in C_i$ are consumption sequences, with

$c_{it} \in R_+^U$ $(t \geq 1)$ being consumption at date t.

$\nu_{io} \in R_+^{I+J}$ $(\nu_{ioi} = 1, \nu_{ioj} = 0, j = 1 \ldots (i\text{-}1)\,(i+1) \ldots I,$

$0 \leq \nu_{ioj} \leq 1, j = (I+1) \ldots (I+J))$ denotes initial shares for

household i in holding companies and in production firms,

respectively. c_{ibo} denotes initial bond holdings. $v_i = (v_{it})$,

$\nu_{it} = (\nu_{itj})$ and $c_{ib} = (c_{ibt})$ are defined correspondingly.

The household in this model thus has no initial endowments. Its initial wealth is entirely of a financial nature, consisting of shares and bonds, while all physical capital is held and managed by the holding companies. Its disposible income in consecutive periods will therefore be made up of dividends from shares held and interest on bond holdings. For simplicity's sake we further assume for the moment that there are no durable consumption goods, so that the household's current purchases are fully consumed in each period.

Let us denote dividends in the various firms and companies at time t, $d_t = (d_{it})$ $(j = 1 \ldots (I+J))$

and represent the value of all shares in the respective firms at the same time $v_t = (v_{jt})$ $(j = 1 \dots (I+J))$.

The budget restriction of household i can then be written as:

$$(2.1) \quad p_t c_{it}^{x)} + c_{ibt} + v_t(\nu_{it} - \nu_{i,t-1}) \leqslant (1 + r_{t-1})\, c_{ib,t-1} + \nu_{i,t-1} d_t$$

The left-hand side of (2.1) consists of the different forms of household outlays - consumption purchases, bond purchases and net purchases of shares. The right-hand side contains two forms of household income - interest on bond holdings and dividends.

Each household i is assumed to maximize $\gtrsim_i$ on the set of all c_i, c_{ib} and ν_i that fullfill the budget restrictions with given c_{ibo} and ν_{io}.

For the production firm j we assume that its production possibilities in any period t can be represented by a set $(z_{jt}, q_{j,t+1}) \subset R^U \times R^U$ where z_{jt} is <u>input</u> at <u>date t</u> and $q_{j,t+1}$ is <u>output</u> at <u>date t+1.</u> We thus assume that output in any period depends on inputs in the preceding period but not on inputs at earlier dates.

Each period the production firm rents its inputs from the holding companies in the form of services or use of material goods and non-producible factors held by these companies. The firm pays, with interest, for the inputs advanced by the holding companies when these inputs have matured after one period into outputs. The firm is thus free from both financial worries and responsibility for the management of physical capital, since these functions are handled separately by the holding companies. These can therefore be looked upon as representing the pure capitalist functions while the production firms account only for current production decisions, i.e. handle the production technology. The outputs of the firms are sold to households for consumption or to holding companies for what we might call gross investment. The dividend at time t will be:

$$(2.2) \quad d_{jt} = p_t q_{jt} - (1+r_{t-1}) p_{t-1} z_{j,t-1} \quad (j = (I+1) \dots (I+J))$$

The firm is supposed to maximize the discounted value of future dividends, where the discount factors are defined as:

$$(2.3) \quad \beta_1 = 1$$

$$\beta_t = \frac{1}{1 + r_{t-1}} \quad (t \geqslant 2)$$

---

x) For a further development of this term, see footnote in the beginning of section 1.5 below.

From our assumptions above, however, it is apparent that, since each dividend depends only on current technology, prices and interest, this is equivalent to letting the firms maximize each consecutive dividend separately, i.e. having them act in a myopic fashion.

For the starting period we must assume that inputs in the previous period, z_{jo}, are given. With this restriction, the behavior rule for the firms can be stated simply. In each period t each firm maximizes its dividend as given by (2.2) above on the set of its production possibilities $(z_{jt-1}, q_{j,t})$.

1.3 The holding companies

In most standard models of this kind only two types of institutions are recognized: households and firms. Between them they then share the responsibility for managing the social capital. The household may e.g. be the owner not only of its own labor capacity but also of land resources, i.e. the households hold the nonproducible resources, while the firms hold the produced capital.

In this model we have chosen to make a further division. Mainly in order to afford human capital a treatment analogous to that of physical capital, we have separated the resource management function into a separate institution called a holding company. Each household is assumed to have its own holding company, which manages various amounts and sorts of the three different kinds of social capital: real estate, human resources and production capital, i.e. reproducible capital including everything from raw materials to durable machinery. What then remains for the household proper is mainly a consumption-saving function although we have found it convenient for later use of the model to let the household also retain the financial investment function, i.e. to be able to buy not only savings bonds but also shares in production firms. The function remaining for the production firms is the economic use of a given production technology.

The holding company owns and manages the human resources - the human capital - of the household. This arrangement may sound strange but means merely that for accounting reasons we have chosen to separate the resource managing functions of the household. This may be expressed by assuming that the current trend towards conducting household business through companies - mostly due in fact to tax reasons - has become the rule.

Each holding company finances its operations for each period separately by selling bonds to households on the bond-market. With this money it purchases the various kinds of non-human capital in the beginning of each period. It then immediately rents out its capital to production firms. In the special case of human capital the holding company also rents part of it - the leisure hours - back to the household at the current market wage. At the end of each period it gets back the capital - or what is left of it - and a rent payment with interest. It then sells out the capital and uses the proceeds together with rent- and interest income to pay back the bond loans, while the surplus is given back as dividends to the household as sole owner of the company.

Since there is no trade in holding company shares there will be no market price. The same is true of human capital, for which a discounted value can be defined, but no market price.

The explication of the human capital concept in general equilibrium conditions would undoubtedly be much simpler and more straightforward if we dared, for modelling purposes, assume some kind of trade in human resources. Before proceeding, it may be worthwhile to comment on the reasons why this does not seem possible.

The reasons are usually regarded as obvious. You cannot go out on the market and buy human beings - there is no slave market. If this argument is taken to imply that labor contracts are always such that they only cover the short-term disposal of labor for specific tasks, then this may not be wholly convincing.[x)]

Any attempt to introduce human capital trade into a general equilibrium model does, however, run into other difficulties which have to do with the individual and indivisible nature of human capital. To account for the fact that you are only interested in buying and consuming your own leisure time, not the time of some equally worthy individual, and that your incentives to utilize training opportunities depend on your reaping the benefits, each person or each household would have to be identified as a separate kind of human capital. Equally obvious is the fact that while you may part-train a man, you cannot train just part of him - from the point of

---

x) Most people are not hired on a purely temporary basis - on "cotton-picking contracts". Employment contracts may assume a long-term view and various employment benefits may be credited to the employer on this assumption. This in fact is what is usually meant when current discussions refer to the trend towards "Japanized labor". A corresponding assumption of long-term utilization can often be traced - and is sometimes explicitly stated - in the conditions for various public employment benefits and training programs. One way of modelling these actual conditions would undoubtedly be to allow for the possibility of selling part of one's own human capital. Contracts can of course always be broken and tacit assumptions can be proved wrong - but usually there is some penalty involved. This then would merely mean - in model terms - that you can always buy back full control of your own labor even if you have partially "hocked" it, as it were, to your employer or some public agency - but this usually involves raising some extra money to make up the part of your capital value you had already mortgaged.

view of developing human capabilities, man is indivisible. If you therefore sell part of him, any change in his capabilities occasioned by his use will introduce externalities between the various agents involved. It thus seems that to assume trade in human capital in an attempt to model real life conditions creates more problem than it solves.

In terms of the kind of generalized activity analysis used here, the treatment of durable capital presents no special problem as long as capital ownership is kept within the production firms themselves. The capital goods before usage are accounted for as inputs and the used and possibly also changed capital goods are registered as outputs in the production technology. There is then no need for explicitly accounting for either use or the accompaning change brought about in the capital goods. When ownership and resource management is placed outside the production firms as in our model here, a new need arises to identify the flow of services between the resource users - the firms - and the resource owners - the holding companies. We have to consider not only the fact that the resources are used but also how they are used, since this affects the shape of the resources when they come back from their use in production. One reason for using holding companies in the model is indeed to make room for this kind of separate service accounting. When it comes to human capital we are not only concerned with what it can do in production, but almost equally interested in identifying what various productive processes can do to develop human capabilities.

The process of accounting for service flows can be modelled in various way. We have chosen to assume that two identifiable "capital services" are involved in any use of a capital resource of a certain kind and quality in production. One is the resource user service, an input in production, which depends only on the resource hired, not on the purpose for which the resource is rented. The user charges can be assumed to be positive throughout. The other is the resource development service, an output in production which, like any other output, depends on the kind of productive activity carried on. The development charges may be either positive - reflecting an improvement in the resource through production usage - or negative - compensating a depreciation in the capital good. The sum of user charges and development charges is the total rental paid for the capital good. For production capital we have made the simplifying assumption that these goods are affected in an identical way by all productive uses, which means that only one rental service and one positive, rental price have to be identified.

The holding companies' transformation possibilities thus in general involve transforming a capital good and a development service into another capital good and a user service. As long as there are a finite number of ways in which a capital good can be changed, the activity analysis can obviously always, in a formal sense, be transformed into this kind of description by a suitable increase in the still finite number of goods in the model. Whether it is also a practically convenient way of modelling real life will depend on the number or standardization of the development services for human and real estate resources that production gives rise to.

It is natural although by no means necessary to think of the transformation possibilities for a holding company as additive, i.e. as being the sum of the separate possibilities for various kinds of capital goods. This would mean that there are no "external effects" between different kinds of resource management. If the transformation possibilities concerning real estate and production capital differ between households this must be interpreted as due to differences and gaps in the mercantile knowledge necessary for handling special kinds of resources. Some households may e.g. not be equipped to handle certain complex kinds of real estate.

We also assume production technology to be such that the volume and character of development services produced with certain resources can be varied and furthermore that these services can be assigned in various ways to the resource units involved in the production process. Specifically we assume that for each relevant kind of resource there is one productive employment actually used which leaves the resource - real estate or human - unchanged.

The household, however, must buy the development service from the same firm which utilizes the unit to be developed. This kind of "tied sales" usually introduces an element of arbitrariness into market pricing. Under the above-mentioned assumptions, however, this will not be the case. The working of the model will be the same regardless of whether we account for the two kinds of services separately or not. There will always be one unique way of splitting the total rental into a user charge and a development charge.

After these introductary remarks we can continue specifying the model by dealing in turn with the three factors of production handled by the holding companies.

1.4 Factors of production

There are three kinds of factors of production in the model: real estate, human resources and production capital.

Real estate in this context refers to more than just land. It covers everything implied by the French word "immobilier", i.e. it includes all immovable properties or objects on land, e.g. roads, various cultivations and land improvements, houses and fixed machinery. The use of such a broad definition of real estate is needed once we want to remove the ownership of the factors of production from the production firms and establish instead a short-term leasing market. There is e.g. no acceptable way of establishing housing rents without involving land use, etc.

A piece of real estate can be described in terms of two different kinds of characteristics, unchangable and changable, respectively. Unchangable characteristics are the various locational properties, geographical coordinates, geological and climatological conditions, etc., that is, everything that determines its latent use and development possibilities. We assume that all existing real estate can be partitioned into a finite number of groups or types, each homogeneous as to these unchangable characteristics, and that the number of units in each such type is given once and for all. Real estate is thus non-reproducible - and eternal - as far as type is concerned.

Through various kinds of land improvements and construction activities the real estate acquires new changable characteristics, which together define the actual state of a certain piece of real estate. We assume here that there are a finite number of such possible actual states and that current real estate rentals depend only on these states. Real estate is thus in a certain sense reproducible when it comes to the actual state of the land. A piece of land will over time pass through a certain state-cycle. The responsable holding company can, by buying different real estate development services, determine each step in this cycle.

Each unit of real estate belonging to a certain holding company at time t can thus be identified as, $e_{it}^{\alpha\beta}$, where α gives the type while β in a corresponding manner refers to the state. We will let $e_{it} \varepsilon R_t^U$ denote a vector with zero for those components that do not correspond to some particular kind of real estate owned by the i'th holding company at time t. By $e'_{it} \varepsilon R_t^U$ we denote correspondingly the one period use of real estate in different states. Given the assumptions made, the number of non-zero components in e'_{it} will at most equal the number of different states.

Finally we use $\dot{e}_{jt} \varepsilon R_t^U$ to denote the real estate development services rendered at time t by the j'th production firm. The number of possible services of this kind - forms of construction, land improvements, etc. - is assumed finite but may be smaller, or bigger, than the number of different kinds of real estate.

One consequence of our definition above may be worth pointing out. In removing all "immovables" to outside ownership we have also eliminated the most common explanation for differences in production possibilities between firms. The task of these firms is now restricted to combining current available services in the most profitable way. The remaining differences in technological possibilities in a certain period must then wholly be ascribed to differences in technological knowledge between the firms. However, the change is more semantic than factual. Instead of saying that a firm can use certain processes because it already has a certain plant built we now say that although it does not itself own the plant it is the one who knows how to operate it.

When we pass to the second factor, human resources, we use definitions and concepts, analogous to those utilized above for real estate. There is one aspect of human resources that is especially troublesome in a model with infinite horizon. People, in contrast to land, die sooner or later. To avoid the complexities arising from this, we make the following simplifying assumptions. Each household reproduces itself indefinitely in such a way that its size and structure remains unchanged. The household may therefore calculate as if each individual of a certain type, i.e. with certain unchangable or inate characteristics, was immortal but necessarily passing through a predetermined life-cycle of aging before "passing into its second childhood".

By various kinds of training, i.e. by being exposed to various kinds of human development services in production firms, an individual may acquire different sorts of skills and capabilities, new states, and thereby pass into new phases of a state-cycle.

Analogous to what was said above about real estate, we assume that these possibilities for transformation or cycling depend on the type or inate characteristics.x) While types of individuals are non-reproducible, states are in this sense reproducible. The model thus encompasses all sorts of "life-long education", while all training is defined as "training on the job", even for the case where the job is just training.

Corresponding to our real estate definitions, we let $h_{it} \varepsilon R_t^U$ denote the vector measuring the human resources of the i'th holding company at time t, with non-zero components indicating its holdings of the various combinations of type and state of labor.

Likewise we let $h'_{it} \varepsilon R_t^U$ denote the use of the different states of human resources, assuming again that the number of non-zero components will at most equal the number of different states. While the individual type determines his development possibilities we thus let his current usefulness in production depend solely on his state, i.e. his actual skill or capabilities.

Finally $\dot{h}_{jt} \varepsilon R_t^U$ represents the various human development services or training (or detraining) opportunities produced at time t by the j'th firm. Nothing is assumed about the number of such opportunities except that it is finite.

---

x) A similar treatment of training as being produced jointly with commodities and "sold" jointly with employment, has been presented by Rosen (1972). He starts off, however, at the point where this paper ends i.e. with homogeneous human capital. Since his purpose is rather to explore some implications of "training on the job" for the choices of employers and employees, his concepts are furthermore not framed within a general equilibrium context. Our MGE-model would seem to meet - and pass beyond - the suggestions for generalization of the Rosen-model put forward by Rosen himself in his note 9.

The third kind of factor is <u>production capital</u>, i.e. all reproducible inputs that are not incorporated into real estate or human resources. Although even in this instance one may talk about recycling in the sense used in industrial and environmental economics, there are no longer a limited number of non-reproducible individuals of certain given types. For simplifying reasons we assume here that for each commodity there is just one standardized change brought about by its use in any production process. To each unit of a capital item thus corresponds a unit of one capital service. Analogous to the definitions above, we denote the production capital owned by the i'th holding company at time t by $\bar{q}_{it} \in R^U_t$, and the <u>production capital services</u> produced by the i'th company $q_{xt} \in R^U_t$.

For "current inputs" or "circulating capital" the transformation possibilities for the holding companies will show the capital services only as output, the capital having been consumed in the process. Should one also wish to allow for "consumer's durable", this simply means that production capital services can be bought from the holding companies also by households.

1.5 Completing the specification

As a general representation of the capital goods bought as input by the holding company i at time t, we use k^-_{it}, and for the corresponding output k^+_{it}, defined as follows:

(5.1) $$k^-_{it} = e^-_{it} + h^-_{it} + \bar{q}^-_{it}$$

(5.2) $$k^+_{it} = e^+_{it} + h^+_{it} + \bar{q}^+_{it}$$

In the same way we define the capital use services produced by the holding company as k'_{it} :

(5.3) $$k'_{it} = e'_{it} + h'_{it} + \bar{q}'_{it}$$

The only difference between the capital use services produced and those consumed by the production firms is the consumption of leisure time and of real estate services by the household. We denote these by c_{ist}:x)

---

x) To take into account the household's purchase of leisure and of real estate services, to be paid afterwards like all capital services, the budget restriction for the household given above in (2.1) must be rewritten as:

(2.1b) $$p_t c_{igt} + (1+r_{t-1}) p_{t-1} c_{is,t-1} + c_{ibt} + v_t (\nu_{it} - \nu_{i,t-1}) \leq$$
$$\leq (1+r_{t-1}) c_{ib,t-1} + \nu_{i,t-1} d_t$$

where $c_{igt} = c_{it} - c_{ist}$, i.e. all consumption except capital services.

(5.4) $\sum_i (k'_{it} - c_{ist}) = \sum_j z_{jt}$

Capital development services were defined separately only for real estate and human resources, which leads to the following definition for the purchases of the i'th company:

(5.5) $\dot{k}_{it} = \dot{e}_{it} + \dot{h}_{it}$

The utilization or transformation possibilities for holding company i in period t can then be represented by set

$T \subset R^U \times R^U \times R^U \times R^U$ where $(k^-_{it}, k'_{it}, \dot{k}_{i,t+1}, k_{t,t+1}) \in T$.

k^-_{it}, and $\dot{k}_{i,t+1}$, are both inputs although one period apart and the outputs, k'_{it}, and $k_{t,t+1}$, are separated in the same way.[x)]

Apart from transformation possibilities for all periods the holding company i will also be characterized by the initial capital, capital leasing and debt, i.e. k_{io}, k'_{io} and k_{ibo}.

The dividend for the holding company i at time t is defined so as to include the imputed value of leisure time. It can then be written as:

(5.6) $$d_{it} = (1 + r_{t-1})p_t\ k'_{i,t-1} + p_t(e^+_{it} + \bar{q}^+_{it}) + k_{ibt} - p_t\dot{k}_{it} -$$

$$- p_t(e^-_{it} + \bar{q}^-_{it}) - (1 + r_{t-1})k_{ib,t-1}$$

The first three terms on the right-hand side represent the

---

x) The representation of transformation possibilities is simplified here by not explicitly stating the restriction to tied purchases of development and user services.

The net product of the holding company at time t can be written as: $k^+_{it} + k'_{it} - \dot{k}_{it} - k^-_{it}$

If this net product is integrated over the whole economy with the corresponding net product for the production firms, the flows between companies and firms will no longer appear in the final term. Apart from household consumption including leisure time, the total net product will evidently only show the difference

$$\sum_i (k^+_{it} - k^-_{it})$$

denoting the demand for investment in new capital goods. In such an integrated model we no longer get a separate representation of the demand for investment in old capital goods, $\sum_i \dot{k}_{it}$.

company's sources of income, i.e. user charges with interest, sales of capital goods and new bonded loans. The last three terms correspondingly denote the various outlays, i.e. development charges, capital goods purchases and repayment with interest of old bond loans.

The holding company will try to find such time sequences of

e^-_{it}, $\bar{q}^-_{it}$, $\dot{k}_{it}$, k'_{it} and k_{ibt}

that will <u>maximize the discounted value of expected future dividends</u>, $\sum_{t=1}^{\infty} \beta_t d_{it}$, with the restrictions given by the period transformation possibilities and the initial values.

If we utilize the assumption mentioned earlier of additive transformation possibilities, dividends could be defined separately and maximized for each of the three sections within the holding company: the real estate department, the labor department and the material department. Since operations in real estate and material are financed separately by bond loans in each period and the transformation possibilities at each time t depend only on the decisions taken at the preceding time, $t - 1$, we immediately see that the myopic quality, which was found to characterize decisions in the production firms, would also be true of operations in real estate and material. Each of these departments could equally well maximize its dividends separately for each period. Since we do not have trade in human capital, this possibility of myopic decision-making does not exist where human resources are concerned. The company must then itself take into account the effects of current decisions concerning employment and training on future labor dividends, instead of having this done by the capital evaluation of a market.

To complete the specification of the equilibrium model, all we need now are the equilibrium conditions. There are three of them. For each $t \geqslant 1$ the following should hold:

$$(5.7) \quad \sum_j q_{jt} + \sum_i (e^+_{it} + \bar{q}^+_{it}) + \sum_i k'_{it} = \sum_i c_{it} + \sum_i (e^-_{it} + \bar{q}^-_{it}) + \sum_j z_{jt} + \sum_i \dot{k}_{it}$$

$$(5.8) \quad \sum_i c_{ibt} = \sum_i k_{ibt}$$

$$(5.9) \quad \sum_i \nu_{ijt} = 1$$

The first equilibrium condition, (5.7), establishes equilibrium on all non-financial markets. The three kinds of supply on the left-hand side are, respectively: production of firms, capital goods sales and capital leasing by companies. On the right-hand side are the four kinds of corresponding demand: household consumption including leisure time, company purchases of real estate and material goods, firm purchases of inputs and finally company purchases of development services.

The second condition, (5.8), simply states that supply should equal demand also in the bond-market.

(5.9) states that conditions on the share market are such that for every share of every firm there is always someone willing to hold it at the going price.

This ends the specification of the model. Apart from stating the equilibrium conditions we have characterized and stated behavior rules for the three kinds of agents, consumers, firms and companies.

Perhaps a word should be added on the v_{jt}, the value of the j'th firm's stock of shares at time t, since the determination of these market values may not be apparent from the specification. In equilibrium, any change of stock value over a period can only be due to the expected dividend. This means that if we assume, as we already have done above, that the discounted value of the sum of future dividends converges, then the value it will converge towards is the discounted stock value, i.e.:

$$\beta_t v_{jt} = \sum_{\tau=t}^{\infty} \beta_{\tau+1} d_{j\tau+1}$$

A picture of the structure of the assembled model for period t with aggregated sectors is presented in fig. 1. For the production firms collectively and individually the choice in each period is, as shown, simply that of picking an input-output pair that maximizes the dividend. The households taken together must distribute their dividend income in an optimal fashion between consumption and bond investments. The task for the holding companies, the truely "capitalistic" task, is somewhat more complex. They must find combinations of on one hand capital utilization and on the other hand investments in new capital goods or in existing real estate and human capital, that will maximize the discounted value of future dividends.

1.6 Income, wealth and the rate of return on human investments

Within the model, as specified above, some concepts of social accounting can easily be defined.

Gross national product (GNP) in period t can be directly defined as:

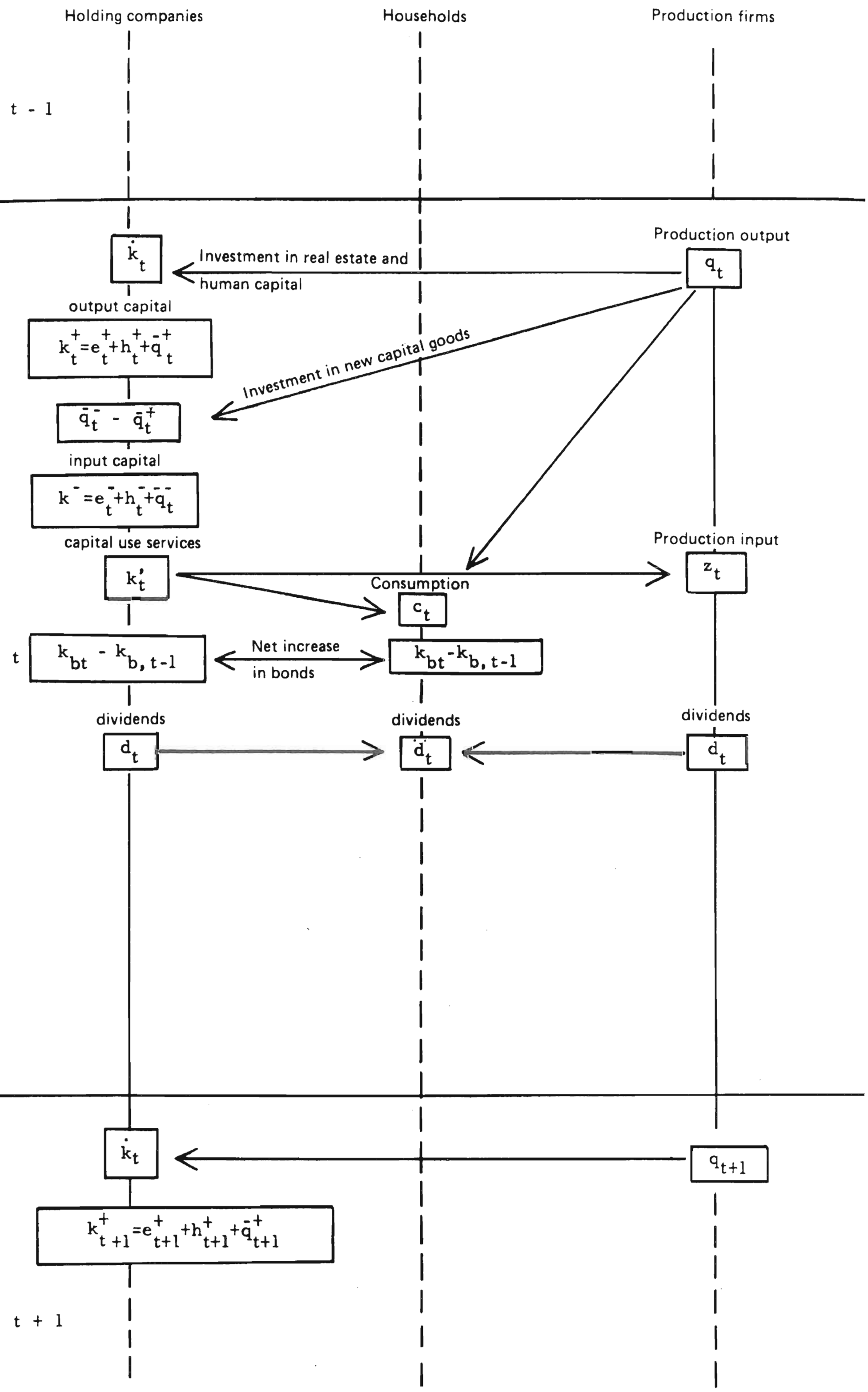
$$(6.1) \qquad GNP_t = \sum_i p_t\{\dot{k}_{it} + (\bar{q}^-_{it} - \bar{q}^+_{it})\} + \sum_i p_t c_{it}$$

This simply means that:

GNP_t = Gross Investment + Consumption

There are two kinds of investments involved. The first term on the right-hand side of (6.1) denotes the investments in non-

Figure 1. The structure of the model in period t with aggregated sectors



reproducible factors, i.e. the total value of development services. The second term represents the gross investments in new material goods and reproducible factors. The third term, denoting consumption, also includes consumption of leisure time. Although this is against social accounting practice, it is motivated here by the importance of leisure value as part of the yield of human capital.

Viewed from the production side the GNP_t could equally well be written as:

$$(6.2) \qquad GNP_t = \sum_j q_{jt} + \sum_i c_{ist}$$

where $c_{ist} = c_{it} - c_{igt}$, as earlier introduced, stands for the household consumption of real estate services and leisure. These services are thus added to the current output of the production firms to give the GNP. In these definitions we abstract from the fact that, in the model, service consumption is paid with interest in the succeeding period, as are service inputs to production. The internal loan transactions between households and their own holding companies are of no concern for the social accounting purposes indicated here.

To go from GNP to national income concepts we have to use capital values to measure changes in wealth. Since there are no market prices for human capital - or formally since p_t is defined with zero components for human capital - we must somehow directly define <u>human capital values</u>. This can be done analogous to the way stock values of firms were defined (see section 5 above), i.e. as discounted values of future expected ·lds.

Let us make the following simplifying assumptions. We assume that the transformation possibilities of the holding firms are additive also down to each individual unit of human capital, i.e. each labor unit can equally well be managed and accounted for separately. This means that for each individual at each time, we can identify in equilibrium the most profitable sequence of future yields. We further assume that holding companies have equal opportunities when it comes to managing labor. Since holding companies are managed competitively, maximizing the discounted value of the sequence of future yields, i.e. the capital value of the human resources, this means that in equilibrium each unit of human resource of a certain type and in a certain state will have the same capital value. One such unit, let us call it h^* (a vector with only one non-zero component) will undergo a cycle of transformations over time although retaining its individual identity. Its capital value at time t can be defined as:

$$(6.3) \qquad v_{h^* t} = \frac{1}{\beta_t} \sum_{\tau=t}^{\infty} \beta_{\tau+1} \{ (1 + r_\tau) p_\tau h^{*\prime}_\tau - p_{\tau+1} \dot{h}^*_{\tau+1} \}$$

The expression on the right-hand side of (7.3) is simply the discounted value of the future rentals or yields of the human capital through its various developments, made up of sales of user services, paid with interest after use, and development costs, respectively.

We can now use these imputed human capital values to complement the price vector so as to get an <u>imputed price vector</u>, $\bar{p}_t$, with non-zero prices also for human capital.

With the help of this price vector we can now define a <u>net national income</u> in period t, NNI_t as follows:

$$(6.4) \qquad NNI_t = \sum_j (v_{jt} - v_{jt-1}) + \sum_i (\bar{p}_t \bar{k}_{it} - \bar{p}_{t-1} \bar{k}_{i,t-1}) + \sum_i p_t c_{it}$$

The meaning of this can also be expressed as:

NNI_t = Wealth formation + Consumption

The first term on the right-hand side of (6.4) denotes the change in total stock values of the production firms, while the second represents the corresponding change in value of the total stock of capital goods. This last expression can also be developed in the following way:

$$(6.5) \qquad \sum_i (\bar{p}_t \bar{k}_{it} - \bar{p}_{t-1} \bar{k}_{i,t-1}) = \sum_i \bar{p}_t (\bar{k}_{it} - \bar{k}_{t-1}) + \sum_i (\bar{p}_t - \bar{p}_{t-1}) \bar{k}_{t-1}$$

The wealth formation in capital goods can thus be split up into two parts: <u>change in the physical volume of capital goods</u> and change in capital prices, i.e. what is usually called <u>capital gains</u>.

Contrary to usual accounting practices, the income concept defined here includes wealth formation also in human capital and capital gains for all capital goods. This seems natural when, as here, we are especially interested in tracing the effects of changes in human capital. It means, however, that part of the measured income changes will materialize as actual purchasing power for the households and their holding companies only when future yields are realized and future dividends paid out. The <u>social wealth</u> concept used above, made up of the stock value of firms and the value of capital goods including human capital, can be derived as the discounted value of all future dividends, all future net incomes or all future consumption. This has already been demonstrated for stock values and dividends and has been defined for human capital. That it is also true for other capital goods depends on the fact that their services in the model are assumed to be sold at a given market price, so that any "producer's surplus" will not affect the value of capital goods for the owners, but will instead be registered in the firm dividends. The net national income can therefore alternatively be introduced as the interest on this social wealth.

If we compare the definitions of NNI_t and GNP_t we see that gross investment in the product is replaced by wealth formation in the income concept.

To simplify the subsequent discussion let us assume constant returns to scale in production, which means zero dividends for the production firms. This is indeed a very natural assumption in this model, where the firms have no durable capital of their own.

The difference between gross investment and wealth formation in the holding companies is not entirely due to physical consumption or potential deterioration of capital goods used in production. There are two other factors involved, one of which is capital gains. The other is the fact that when it comes to real estate and human resources there is no given relation between the cost of a certain development service and the value of its impact on the capital employed. Training opportunities offered by a production process may e.g. be of great value to employers of a certain type and state although they are only charged the common market price, which may be insignificant.

We can go on to define gross profits in a similar way. Profits in the production firms will equal dividends, i.e. are zero according to the assumption made earlier. If we consolidate the rest of the economy into a household sector to eliminate loan transactions <u>imputed gross profit</u> for this sector in period t, $\bar{\Pi}_t$, is then defined as:

(6.6) $$\bar{\Pi}_t = \sum_i \{ (1+r_t) p_t k'_{it} - p_{t+1} \dot{k}_{i,t+1} \} + \sum_i \bar{p}_{t+1} k^+_{t+1} - \sum_i \bar{p}_t k^-_t$$

The expression on the right-hand side is made up of three terms, each summed over all households. The first profit term designates rentals, i.e. the net payment received from the firms for services plus the value of household consumption of leisure and real estate services. The second is the value of output capital and the third denotes the value of input capital. The word "imputed" refers to the fact that imputed value changes in human capital have also been included. If we take into account the interest cost involved in holding the input capital over one period we get the <u>imputed net profit</u>, $\bar{\pi}_t$:

(6.7) $$\bar{\pi}_t = \bar{\Pi}_t - r_t \sum_i \bar{p}_t k^-_t$$

The <u>imputed gross profit rate</u> in period t, ρ_t, can be written as:

(6.8) $$\rho_t = \frac{\bar{\Pi}_t}{\sum_i \bar{p}_t k^-_t} = \frac{r_t \sum_i \bar{p}_t k^-_t + \bar{\pi}_t}{\sum_i \bar{p}_t k^-_t}$$

It follows directly from this definition that if net imputed profit is zero, $\rho_t = r_t$.

Applied separately to human capital this simply expresses the trivial relation that if income from human resources including increases in capital values is just enough to cover current interest on incoming human capital then the imputed gross profit rate for human capital will also be equal to this current rate of interest.

The concepts discussed so far are all total or average. Marginal concepts can be dealt with more conveniently if we assume for the time being that the transformation possibilities of the holding companies can be expressed by differentiable functions. If we also add the harmless assumption that capital services supplied have as their unit of measurement the services of one unit of the corresponding capital good, the transformation possibilities can be written as:

(6.9) $$k'_t = Mk^-_t$$

(6.10) $$k^+_{t+1} = f(\dot{k}_{t+1}, k^-_t)$$

(6.9) states that user services supplied in period t constitute a linear function of incoming stocks of capital goods. The quantity of user services of a capital good in state β is simply the total number of capital goods in this state summed over all types α. M here thus stands for a quadratic matrice of order U x U, where each row corresponding to a component of k'_t measuring user services of goods of a certain state has 1:s in all the places which measure capital goods of this state, while the rest of the matrice is made up of zeros.

(6.10) says that outgoing capital is a function of incoming capital and development services. Both these functions are then assumed to be differentiable. This is certainly not strictly the case, since it is a fundamental property of capital goods as defined here - and particularly of human beings - that they can only change as a unit. Our excuse for using this assumption here is the usual one employed by economists; aggregates may be taken to be large enough and, anyhow, discontinuities do not seem pertinent to the analysis.

We also abstract from the fact that the second function can be multi-valued due to the possibilities of different assignments of development services to the various capital units, i.e. we assume a given assignment rule that makes it possible to trace an expansion in a unique fashion.

If we differentiate (6.9) and (6.10) we get:

(6.11) $$dk'_t = Mdk^-_t .$$

(6.12) $$dk^+_{t+1} = f_{\dot{k}}\, d\dot{k}_{t+1} + f_k dk^-_t$$

If we take the price vectors $\bar{p}_t$ and $\bar{p}_{t+1}$ as given there are then two possibilities of marginally affecting profits and profit rates - by changing incoming capital and by changing the investment in development of real estate and human resources. Let us look at these in turn.

Marginal changes in incoming capital cannot be made in human resources, which at each point in time are given and non-tradable. With this restriction in mind we can define the marginal change in imputed gross profit resulting from a marginal change in incoming capital as:

$$(6.13) \qquad d\bar{\Pi}_t = \sum_i \left[(1+r_t)\, p_t\, M + (f_k \bar{p}_{t+1} - \bar{p}_t) \right] dk_t^-$$

If no development service is reassigned to the new capital we can neglect f_k and gross profits will change due to change in user charges and in the volume of capital gains. But we also know from differentiating (6.7) that:

$$(6.14) \qquad d\dot{\Pi}_t = d\bar{\pi}_t + r_t \sum_i \bar{p}_t\, dk_t^-$$

Since we are studying an equilibrium point where net profits will be maximized along with dividends, marginal net profit is zero which obviously means that the marginal profit rate will equal the rate of interest. This is what we usually expect to find in equilibrium models, that is:

$$(6.15) \qquad d\rho_t = \frac{d\bar{\Pi}_t}{\sum_i \bar{p}_t\, dk_t^-} = r_t$$

If instead we study a change in development investments in real estate and human resources, the corresponding change in imputed gross profits will be:

$$(6.16) \qquad d\bar{\Pi}_t = \sum_i \{ \bar{p}_{t+1} f_k' - p_{t+1} \}\, d\dot{k}_{i,t+1} =$$

$$= \sum_i (\bar{p}_{t+1}\, dk^+_{i,t+1} - p_{t+1}\, d\dot{k}_{i,t+1}) = 0$$

That the change in profit must, in equilibrium, equal zero follows from (6.14) above. Not only must marginal net profit again be zero but this is now also true of the second term in (6.14), the change in interest on incoming capital. Applied to human investments, (6.16) states that the value of a marginal investment in equilibrium will equal its cost. It follows that the marginal profit rate is also zero:

$$(6.17) \qquad \frac{d\bar{\Pi}_t}{\sum_i p_{t+1}\, d\dot{k}_{i,t+1}} = 0$$

This may at first seem more surprising, as it means that e.g.

the marginal rate of profit on investments in human capital is zero. This, however, is merely a consequence of the fact that in the model we assume such investments to change the human capital immediately without any costly delays. Here we have also let these investments affect the imputed gross profit directly by way of the imputed capitalization of future increases in human rentals. If we rewrite (6.16), substituting from (6.3) above the full definition of imputed capital value, we get:

$$(6.18) \qquad \sum_i p_{t+1} d\dot{h}_{i,t+1} = \sum_i d\overset{+}{h}_{i,t+1} \{ \frac{1}{\beta_{t+1}} \sum_{\tau=t}^{\infty} \beta_{\tau+1} (1+r_\tau) p_\tau h_\tau^{*\prime} - p_{\tau+1} \dot{h}^{*}_{\tau+1} \}$$

where h^* as before traces the various states of each original unit of human capital. The expression within brackets on the right-hand side simply measures the discounted value of future human rentals or, what amounts to the same thing, the discounted value of future profits on human capital - without imputation of human capital values. The meaning of (6.18) can therefore be written simply as:

The cost of a marginal human investment =

The discounted value of the future marginal profits which result.

The rate of return of marginal human investments, defined as the internal yield rate, will then in equilibrium depend on the spacing of future marginal profits and on the sequence of interest rates involved. Specifically, if the rate of interest remains constant over the future, the marginal rate of return on human investments will equal this interest rate just as it will for other kinds of investment; see (6.15) above.

2 Homogeneity of human capital

2.1 The need for further assumptions

What we have shown in the first chapter is simply that the human capital concept can be incorporated in a general equlibrium context in a way similar to what is done with other factors of production and that it can also be fitted into a social accounting matrix.

We have traced the accounting relations between the financial concept of human capital value and the rentals or earnings of human resources. Human capital in this sense may be a conveniance for social accounting and a shorthand notion for discussing individual expectations.

The reason for using a "human capital approach", however, is usually much more ambitious. One hopes in this way to arrive at some testable relation between the volume of human investments in an individual and his or her earnings, possibly with some capability factor as an intermediate variable.

From this point of view our general equilibrium model is far too general. All we can generally say about the wage of an individual, i.e. his user charge in the model, is that it depends on his state, which in turn depends both on what type of person he is and on how much has been invested in him - assuming that the investments have been optimal. His human value is thus partly accounted for by investments done and partly by his type rent, i.e. the value of his development potential. In principle we could try to separate the two components by measuring at birth the discounted value of all future rentals until death - his individual type rent value which then also incorporates various surpluses in human investments. But we cannot measure either total rent or differential rent separately later on in life. Since people are assumed to be fundamentally different - to belong to different types and not just more or less capable - they will usually be found chosing different careers. The relative remuneration of these careers may well change with demand and supply conditions over time. In the same way the relative costs of various forms of human investments will vary and will not generally be proportionate to their relative value. In one respect, however, the model is better tailored to the needs of human capital theories than real life. In the model pure wages and human investments are accounted for separately, while in real life, we can usually only determine the total rentals for various kinds of jobs.

In order to derive the kind of simple relations between the volume of human capital in some sense and the individual earnings that human capital theories often aim at, the model must obviously be restricted and further specified in several ways.

By stepwise restricting and modifying our model, we will in the following try to approach the kind of human capital theory model needed to "explain" the relative structure of current earnings.

2.2 The first assumption: Physical human capital and the possibility of aggregation

One main assumption inherent in a "pure" human capital theory is that relative earnings depend only on the physical capital, the "earning potential", embodied in the individuals. Let us for the time being keep this assumption as general as possible, still allowing e.g. for a possible heterogeneity of this physical capital. If we call the price component for the i'th state of human capital w_i, which then measures the wage or user charge of an individual in state i, the assumption could be formally expressed as:

$$(2.1) \qquad \frac{w_i}{w_j} = f(h)$$

where we for notational conveniance assume that the relation is a differentiable function.

What (2.1) states is then that the relative wage is a function only of the vector of human resources used. The relative wage is thus unaffected by changes e.g. in production capital, in cooperating real estate, in output mix or in the rest of the equilibrium price vector.

Let us reflect for a moment on what this requirement means in terms of our MGE-model. There we had human resources of various types and states and we assumed that earnings were only related to states. To fit into the assumption discussed here we must then first of all accept that all human resources in the same state, regardless of type, embody the same physical human capital. The distribution between states of embodied capital then determines relative earnings.

Even if the human capital is heterogeneous we cannot explain changes in relative earnings by reference to changes in relative capital prices, as we do in dealing e.g. with machines. There are no capital markets for human resources and any attempt to bring in changing imputed capital prices for various kinds of labor would obviously rob the physical,human capital concept of any explanatory power and bring us back to the starting point, i.e. to the financial concept of human capital value as a discounted sum of future earnings.

The quotient of wages in (2.1) expresses in equilibrium the marginal rate of substitution between the two kinds of factors, i.e. between human resource units of states i and j. The assumption of (2.1) can therefore equally well be expressed in the following manner. The marginal rate of substitution between two kinds of labor should be independent of everything except the amounts of labor inputs.

If we simplify the model to account only for a vector of production capital q, two kinds of labor inputs h_i and h_j and a homogeneous output C, the condition can be more directly stated as follows. "The marginal rate of substitution between the two labor inputs is independent of the output level and the vector of production capital".

In this form the condition is known as the Leontief condition (Leontief, 1947) and expresses a necessary condition for aggregating the two inputs. We can then call our condition above a "generalized Leontief condition" and have as an hypothesis that this condition also expresses a necessary aggregation condition. Let us try briefly to follow up this idea.

In our model we have represented production possibilities as point sets. Let us for illustrative purposes use a simplified form of such a set T with elements of the form (h, q, c) where h still stands for a vector of labor inputs, q for a vector of production capital and c for a vector of output goods.

Aggregation of h in terms of such a model simply means that there exists a well-defined function ϕ, which relates h to a scalar, an aggregate, H: $H = \phi(h)$, and which possesses the following property:

There exists a set T_a with elements (H, q, c), such that $(H, q, c) \in T_a$ if and only if $(h, q, c) \in T$.

(For a more extensive discussion of meanings of aggregation the reader is refered to Fisher (1965), Morishima (1961) and Bliss (1975)).[x)]

That the Leontief condition is a necessary condition for aggregation can easily be seen in the following way. Let us write the differentiable production function as: $C = P(q, h)$. If we

x) The meaning of this can perhaps be better grasped intuitively if expressed in an alternative way.

For every possible collection of production capital, q, and output, c, in T there is paired a set of possible combinations of labor inputs, $\{h \mid (h, q, c) \in T = S_h$. Seeking an aggregating function then means we are looking for a way of ordering these sets S_h in a linear way so that each set can be assigned a unique number H. This is obviously only possible if a complete (and continuous) order is already established between the sets by inclusion, i.e. if the sets are such that either $S_{h_i} \subseteq S_{h_j}$ or $S_{h_i} \supseteq S_{h_j}$. A collection of sets S_h ordered in this way is said to be nested.

The definition of aggregation given above can also be shown to be equivalent to the following condition (for proof cf e.g. Bliss (1975)):

The sets S_h are nested.

substitute the aggregate $H = \phi(h)$ in this function we get $C = P(q, H)$. The marginal rate of substitution between two kinds of labor becomes:

$$(2.2) \qquad \frac{w_i}{w_j} = \frac{P_H \cdot \phi_i}{P_H \cdot \phi_j} = f(h)$$

where P_H and ϕ_i represent partial derivatives with respect to H and an i state component of h, respectively.

As long as we can reformulate our model in terms of differentiable functions this demonstration can obviously easily be extended to encompass our generalized version of the Leontief condition.x)

The concrete meaning of our generalized Leontief condition is easily spelled out. It is obviously a very strong assumption that is only fulfilled in some very special theoretical cases and almost certainly never in real life.

One such case is of course the case of perfectly homogeneous labor, an infinite elasticity of substitution between any two states of human capital.

Another case exemplifies the opposite, with fixed coefficients for different kinds of labor - an elasticity of substitution equal to zero.

A possible third theoretical case would be when labor is already aggregated in real life by being organized in "labor companies" that offer for sale a homogeneous intermediate labor service.

These examples suffice to indicate the strength - and lack of realism - of the assumption.xx)

x) It should perhaps be pointed out that here we have treated aggregation conditions in the most general form, which is motivated by our aim of applying the conclusions also to our very general MGE-model. Had we narrowed our attention only to stationary equilibria the aggregating conditions would have been formulated in a different way, stressing not only the difficulties in establishing a linear order but also the problem of doing so in such a way as to preserve functional relations in the aggregate. This is best exemplified by the discussion, wellknown in capital theory, of the possibility of establishing meaningful chain-indices for capital goods between different stationary equilibria. (See e.g. Champernowne (1953-54) and Bliss (1975)).

xx) If we want to restate the assumptions in terms of a stationary model version (see note at the end of the chapter), it should be observed that the assumption does not require the wage structure to be constant between stationary equilibria. This would be the case if human capital were really "non-reproducible". As it is the wage structure may change between equilibria, not as a direct consequence of other changing equilibrium prices, but indirectly via induced changes in the composition of different states of human capital.

2.3 The second assumption: Homogeneous human capital

The first step towards restricting our model only established that the earning capacity for an individual in a certain state is somehow physically embodied in the individual. This obviously does not take us much further when it comes to finding a uniform explanation for the relative earnings. We have only as it were moved the problem to be explained one step backwards, into the physical character and productivity of the various individuals.

Since it is not the aim of human capital theories to provide explanations of physical differences in productivity between different individuals, the next step to be taken is obvious. We must also assume that human capital is homogeneous, i.e. that any two units of human capital have an infinite elasticity of substitution. This means that the human capital of different individuals, h, can not only be aggregated to H but also measured with the same measuring rod, i.e. in units of H.

When all human capital is one and the same, its distribution between individuals does not matter, by definition (we abstract here from indivisibilities). In (2.1) the wage quotient depended on the variable composition of the human capital stock. With homogeneous capital this functional relation must be constant, determined by the given human capital values, H_i and H_j, of the respective states of human resources.

$$(3.1) \qquad \frac{w_i}{w_j} = f(h) = g\left(\frac{H_i}{H_j}\right)^{x)}$$

We have so far only dealt with what homogeneous human capital is, without explicitly saying anything about how it is formed. Our MGE-model provided a very general description of human investments in terms of "development services", $\dot{h}$. If we assume that the possible transformations of human capital can be represented separately in functional form, the process described in the model for human capital formation can be written as:

$$(3.2) \qquad H_t = \psi^{\alpha}(H_{t-1}, \dot{h}_t) = H_o^{\alpha} + \psi^{\alpha}(\dot{h}_1, \ldots\ldots, \dot{h}_t)$$

x) In studying a linear stationary equilibria we can make this statement more precise. We then know that net rentals on all capital will just pay the interest costs on the capital values. (3.1) thus becomes:

$$\frac{w_i}{w_j} = \frac{raH_i}{raH_j} = \frac{H_i}{H_j}$$

where a represents some given unit price of human capital.

(3.2) says that for an individual of type α his amount of human capital at time t is a function of both his human capital in the preceding period and the development service, the human investment, that he has absorbed since then. This recursive formula can equally well be rewritten, as is done at the end of (3.2), in terms of an initial type capital, H_o^α, and a function of the series of investments up to time t, that measures what we can call the investment-induced capital.

2.4 The third and fourth assumptions: Homogeneity of human investments and the ordering of human types

By means of our two successive restrictions we have now arrived at the concept of a homogeneous human capital, whose actual amount determines the state and through that, also the earnings of an individual. Is this then enough on which to build a general and applicable economic explanation of relative earnings? To see that the answer to this question is still no, it suffices to look at the expression in (3.2.) above.

To determine the amount of human capital and, in turn, the earnings of an individual we must obviously know a) what type of individual he or she is b) what the exact form of the functional relationship in (3.2) is for this type and c) what kind of experiences the individual has had and in what order. This is a rather big order, certainly too big for anyone aspiring to reach a simple and unified explanation in economic terms. Since we assume here that this is the aim of the pure human capital theory we must introduce further restrictions.

One first such restriction must deal with the various kinds of human investments. It does not help us to know that the human capital, which is inaccessible for direct measurement, is homogeneous if there are any number of heterogeneous inputs that can produce this capital. We would then need exact knowledge of the effect of each of these on human capital production. The only assumption that will let us escape from these difficulties is the assumption of homogeneous human investments. We assume in other words that we can treat the $\dot{h}$ as homogeneous, measurable and aggregable in terms of scalar $\dot{H}$, in the same way as we assumed human capital, the h, to be homogeneous. It must be fully homogeneous, not just possible to aggregate in the production function, since we also need to be sure that the order of investments over time can be neglected.

With this assumption (3.2) can be rewritten as:

$$(4.1) \qquad H_t = \psi^\alpha \; (\dot{H}_t)$$

where $\dot{H}_t$ measures the sum of human investments for an individual of type α up until time t.

The usual empirical interpretation of the assumption is of course that the collection of investments in an individual can be measured by the sum of discounted investment costs.

However, there is still a major obstacle left before we can arrive at a unified explanation. As long as the functional relations, ψ^{α}, are left wholly unspecified, we could still have heterogeneous people even though both human capital and human investments are assumed homogeneous. If the set of functions, ψ^{α}, differ between themselves in many ways, e.g. in many functional constants, people of different types will be heterogeneous in that their differences cannot be measured by any one scalar. We would then still have to make specific estimations of the function for each type, which would usually require more information and other kinds of information than we have available to us.

To explicitly exclude this possibility we must assume that the functional relations, the "production functions for human capital", are the same for all types only differing in one "type constant", say α. This means we now assume that (4.1) can be written:

$$(4.2) \qquad H_t = \psi(\dot{H}, \alpha)$$

α can be said to represent some uni-dimensional measure of capability. People may be more or less capable but they are otherwise the same. To give capability an unambiguous meaning one should also add the further assumption that given the same volume of human investments, a more capable person is always a better prospect for further investments. This is analogous to the natural assumption that human investments always have a non-negative yield. These two assumptions may be written as:

$$(4.3) \qquad \psi'_{\alpha} \geq 0$$

$$\psi'_{\dot{H}} \geq 0$$

By means of these last assumptions we have finally restricted our model to the point where it can be said to yield a unified explanation of relative earnings in terms of embodied human capital. "All" we have to do - if we believe in the explanation - is to try to estimate the ψ-function and measure the capability of various individuals on some scalar scale.

We certainly cannot claim that all so-called human capital theories are based on these assumptions. What we can claim is that those theories that have the stated explanatory goal and are conceived within the context of a general equilibrium must make these, or analogous, assumptions. Our guess is that this covers a major part of the literature on human capital theory.

A note on stationary equilibrium

The MGE-model is somewhat unwieldly for serving as an example and frame of reference. Instead of using a general intertemporal equilibrium it is sometimes more convenient to formulate the problems in terms of a comparison between different stationary equilibria, all possible within the given technology but with different equilibrium prices.

This also makes the empirical interpretation easier since in a stationary model you can avoid the difficulty of having to separate wages from net human investments.

In principle you could attain the same advantages by using, instead of a stationary model, a model with semi-stationary growth, i.e. a model where all inputs and outputs keep expanding proportionately at a given rate.

However, this procedure is not open to us since we have explicitly incorporated real estate into our model and have defined types of real estate in terms of given and unchangeable characteristics of land. With our broad interpretation of real estate it would make even less sense than usual to make use of economists' worn-out excuse that land expansion is to be interpreted as a land-augmenting but otherwise neutral technical progress.

Since we are no longer interested in tracing individual investments in human resources we can simplify by reverting to the more traditional institutional arrangements, splitting up the holding companies so that produced capital goods are managed by the production firms while the human resources and real estate are held directly by the households. We may also treat the collection of firms as if it were only one maximizing unit, since the production technologies of the firms add up to the total production plan for the economy, which will maximize profits only if this is true of each component firm.

The equilibrium being stationary there will be no net investment in human resources or in any other kind of resources, which means that in the aggregate, from the firm's point of view, we can treat human resources and real estate as if they were unaffected by production and compensated only by user charges paid when production matures after one period.

For notational convenience we can further partition the price vector p into separate vectors for consumption, and for various factors of production. This means that we treat the same commodity appearing both in consumption and as production input as two separate commodities.

With these modifications the transformation taking place in the firms can then be expressed as the following pair of inputs and outputs, being one of the many possible transformations included in the, once and for all, given set of production possibilities, T:

$$(1) \qquad \{(0, q, e, h), (c, q, 0, 0)\} \in T$$

where q stands for producer inputs, e for real estate, h for human capital and c for consumption.

What (1) says is then simply that, with the help of the "non-producible" factors, real estate and human resources, a given vector of production capital reproduces itself and also leaves a consumption surplus. This is the production cycle that is assumed to repeat itself during each period in the stationary equilibrium.

Another common way of characterizing the stationary equilibrium is to write down the expression for aggregate firm profit. With the use of the partitioned price vector the profit expression will appear in the following well-known format:

$$(2) \qquad \Pi = p_c c - r p_q q - p_a e - p_w h$$

$p_c c$ here represents the income from consumption sales, $rp_q q$ is the interest cost of holding stocks of production capital, while $p_a e$ and $p_w h$ represent rentals for real estate of different kinds and for human resources, respectively, both measured at the time of payment.

This all sounds so familiar that it may be well to remind the reader of some aspects of (2), due to its derivation from the more general model, which make it still somewhat special.

Real estate and human resources here are "non-producible" only as to type but not as to state. Since state is what matters in production, there is still a large, although restricted, choice between alternative distributions between states, reachable within the given technology by way of investments in development.

Secondly we cannot claim here that equilibrium prices will be uniquely determined by the chosen technological transformation. Much of the recent discussion around stationary models has assumed that prices are not dependent on demand and these assumptions have been formalized into various so-called non-substitution theorems. However, there can never be a unique equilibrium price independent of demand if there is joint production of goods that have to face different demands. This is certainly the case in our MGE-model, where development services were specifically defined in terms of a joint production process.

(2) can be still further simplified. Although linearity or constant returns to scale is not a necessary concomitant of stationarity, it is rather hard to find plausible excuses for

not making this assumption. Therefore in the following we assume $\Pi = 0$.

The competitive conditions in equilibrium can now be expressed in a straightforward way.

Since we are dealing with a competitive equilibrium we know that the chosen transformation is more profitable than any other possible transformation at the given equilibrium prices. If we call this transformation number one and number the variables accordingly, we have

(3) $$\Pi^1 = p_c^1 - r^1 p_q^1 q^1 - p_a^1 e^1 - p_w^1 h^1 = 0$$

If we compare this with an alternative transformation called number two, the condition of competitive equilibrium ensures that:

(4) $$p_c^1 c^2 - r^1 p_q^1 q^2 - p_a^1 e^2 - p_w^1 h^2 \leq 0$$

If we subtract (3) from (4) and define $\Delta y = y^2 - y^1$ where y is any vector of values we get:

(5) $$p_c^1 \Delta c - r^1 p_q^1 \Delta q - p_a^1 \Delta c - p_w^1 \Delta h \leq 0$$

Already from this restatement of competitive conditions we notice one main difference from the usual discussions in capital theory. There it is assumed that non-produced factors are strictly non-producible which means that Δe and Δh in (5) are both equal to zero and that we can continue to study in isolation e.g. the impact of capital intensity on consumption standards. When investments in human resources and real estate are brought in as here, the discussion has to be broadened to deal also with changing patterns in these kinds of capital.

When human capital is homogeneous we can use the aggregate directly in reformulating the profit expression for our stationary equilibrium (cf (2) above).

(6) $$p_c c - r p_q q - raH - p_a e = 0$$

If we want to illustrate how human capital theory impinges on the controversial issues in capital theory, this can easily be done with (6) as a starting point.

Let us make all the simplifying assumptions that were used earlier in the more "naive" discussions of capital theory. This means that on top of our earlier assumptions we now also assume that a) consumption and produced capital can also be treated as aggregates and represented by C and Q and b) no land is needed in production, i.e. all components of e are equal to zero. (6) then reduces to:

(7) $C - rbQ - raH = 0$

where b represents the unit price of produced capital. If we now compare two alternative equilibria, using the competitive condition, as in (5) above, we arrive by successive subtracting to the following condition:

(8) $\Delta (rb)\ \Delta Q + \Delta (ra)\ \Delta H \leq 0$

where, as before, $\Delta(rb) = r^2b^2 - r^1b^1$, etc.

In the usual capital theory discussions, labor resources were assumed given and unchangeable, i.e. $\Delta H \equiv 0$. (8) can then be interpreted as expressing the well-known notion that "a stationary state with a higher level of capital relative to non-produced factors cannot have a higher rental of capital in terms of consumption".

Here, with $\Delta H \neq 0$, i.e. with possibilities for human investments, the interpretation of (8) becomes somewhat more complex. The condition now states that "if a stationary state with higher level of produced capital also has a higher rental for this capital, then the resulting cost increase must be at least compensated by either a decrease in human capital or a lowering of the rental on this capital".

Since this whole line of reasoning presupposes uninhibited aggregation throughout it does not really merit much interest. As is well known one outcome of recent controversies in capital theory has been a rather general agreement that an over-all-aggregation of produced capital is never theoretically justifiable. There is an irony in the fact that human capital theory, building on aggregating conditions, matured and was turned into applications at the very time that agreement was reached on the impossibility of capital aggregation in general.

References

Arrow, K.J. (1951): "An Extension of the Basic Theorems of Classical Welfare Economics", Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (ed. Neyman, J.), University of California Press

Arrow, K.J. - Debreu, G. (1954): Existence of Equilibrium for a Competitive Economy, Econometrica 22

Becker, G.S. (1962): "Investment in Human Capital: A Theoretical Analysis", Journal of Political Economy, Supplement, Oct.

Becker, G.S. (1964): "Human Capital", New York

Becker, G.S. (1967): "Human Capital and the Personal Distribution of Income", W.S. Woytinsky Lecture, no 1

Bliss, C.J. (1975): Capital Theory and the Distribution of Income, North-Holland

Champernowne, D.G. (1954): "The production function and the Theory of Capital: a comment", Review of Economic Studies 21

Debreu, G. (1951): "The Coefficient of Resource Utilization", Econometrica 19

Debreu, G. (1959): The Theory of Value, Yale University

Fischer, F.M. (1964): "The Existence of Aggregate Production Functions", Econometrica, no 4

Grandmont, J.M. (1976): "Temporary General Equilibrium Theory", CEPREMAP Working Paper no 7601

Guesneri, R. - Jaffray, J.-Y. (1974): "Optimality of Equilibrium of Plans, Prices and Price Expectation", in Draze, J. ed. Allocation under Uncertainty: Equilibrium and Optimality, MacMillan

Hicks, J. (1939): Value and Capital, Oxford University Press

Koopmans, T.C. (1951): "Analysis of Production as an Efficient Combination of Activities", in Activity Analysis of Production and Allocation (ed. Koopmans, T.C.), New York

Leontief, W.W. (1947): "Introduction to a Theory of the Internal Structure of Functional Relationships", Econometrica 15

Lindahl, E. (1929): "Prisbildningsproblemet från kapitalteoretisk synpunkt", Ekonomisk Tidskrift

Malinvaud, E. (1953): "Capital Accumulation and Efficient Allocation of Resources", Econometrica 21

Malinvaud, E. (1972): Lectures on Microeconomic Theory, North Holland

Mincer, J. (1962): "On-the-job Training: Costs, Returns and Some Implications", Journal of Political Economy, Supplement, Oct.

Mincer, J. (1974): "Schooling, Experience and Earnings", Mimeo, National Bureau of Economic Research, New York

Morishima, M. (1961): "A historical note on Professor Sono's theory of separability", International Economic Review 2

von Neumann, J. (1945): "A model of General Economic Equilibrium" Review of Economic Studies 13

Rosen, S. (1972): "Learning and experience in the Labor Market", Journal of human Resources, vol. 7, no 3

Sato, K. (1975): Production functions and aggregation, Amsterdam

Svensson, L.E. (1976): On Competitive Markets and Intertemporal Resource Allocation, Institute for International Economic Studies

CHAPTER IV

Structural Change as an Equilibrium or Disequilibrium Process – An Introduction

Contents

1. A "Model" History of the Swedish Economy
2. International Trade and Structural Change
3. The Dynamics of Capital Accumulation
4. The Energy Economy
5. Classical versus Keynesian Modelling
6. Market Behavior
7. Government Behavior and Stabilization Regimes
8. Comparing Models
References

1. A "Model" History of the Swedish Economy

The postwar development of the Swedish economy is reflected in the history of its medium-term macroeconomic models. The 50s and 60s were a period of fast growth and relative price stability. Swedish exports were boosted first by reconstruction needs in war-ravished Europe and later by the trade liberalization which, part of the time, was reinforced by an undervaluation of the Swedish crown. The stability of the international monetary system and the seemingly well established pattern of expanding world trade induced the government to take for granted continued industrial competitiveness and the possibility of full employment policies. Medium-term policies were focused on distributing the gains of industrial productivity growth and international trade through a fastly expanding public welfare system.

The main task of the government's medium-term surveys was to check on the consistency of allocative schemes and to provide a framework for the long-term public expenditure plans, which – in the 60s – were becoming increasingly elaborate by means of various PPBS[1]-procedures. When the first medium-term macroeconomic model was developed by the Treasury in the late 60s, its aim was mainly to enhance computational convenience and further ensuring internal consistency in constructing balanced growth scenarios for the medium-term surveys (Åberg 1971). A static multisectoral model was used to compute desired equilibrium values at some future target date and to describe – or prescribe – balanced growth paths from the present to the target. The model only dealt with developments in real terms with consistent aggregate price and wage structures being computed afterwards. Although becoming successively more detailed and elaborate in terms of sector disaggregation etc., the methodology of Treasury models remained unchanged throughout the 70s (Ministry of Finance 1976).

Meanwhile the focus of medium-term problems changed drastically during the 70s. The increased price and exchange rate instability and uncertainty, demonstrated, and partly caused by, the oil price hikes, gave in Sweden rise to dramatic swings in industrial production and investment activity during the latter part of the decade. The shifts in world market demand also revealed a seriously deteriorated position for many of the raw materials and investment goods on which Swedish exports had been traditionally based. Interest became focused on the need fo structural adjustment in industry to eliminate a mounting balance of paymen deficit. The rapid expansion of local government consumption and central govern ment transfers, partly used as a means of "bridging" the employment problem during international recessions, had simultaneously created a domestic budge deficit problem, which, with mounting real interest rates, threatened to lead fisca policy into a "debttrap".

These problems meant new challenges for macromodelling. A major task o medium-term models now had to be to take explicit account of the price uncertain ty and to measure the propagation and impact of changes in world market price

[1] *P*lanning – *P*rogramming – *B*udgeting – *S*ystems.

and in Swedish competitiveness. It thus became of strategic importance to integrate price and wage formation into the models and to be able to use the models for tracing the effects of price changes on industrial profitability and investment in different sectors.

To be able to analyze the mechanisms of structural change the models must furthermore be dynamic in the sense of i.a. taking explicit account of the economic inertia represented by a given vintage structure of industrial capital and the lagged response of consumer demand. Finally, to provide a basis for policy analysis, the models should make it possible to study how various kinds of price policies and stabilization regimes could be used to ease the necessary structural and financial adjustments.

The two models documented in this book represent the first attempts – started in the late 70s – to meet these challenges.[2] They are based on widely different modelling concepts but did nevertheless result from a cooperative effort, incorporate many common dynamic mechanisms, exploit the same data-base and share many common econometric estimates. They were also initially employed to analyze the same problems, namely the impact of oil price shocks on the Swedish economy and possible policy means to insure in advance against, and/or to ease the adjustment after, unexpected dramatic changes in world market prices. The results of that study have already been extensively documented (cf. Ysander 1983b,c).

The first model, ELIAS,[3] is a six sector multiperiod equilibrium model developed by Lars Bergman at the Stockholm School of Business as a dynamic counterpart to an earlier static model of his (Bergman 1982). Like the other model, ISAC,[4] it incorporates price and wage formation and explicit links with the world market, treats capital accumulation in terms of vintages and determines consumer demand by linear expenditure functions. Apart from these dynamic elements it does not however – unlike ISAC – recognize any other sources of market inertia or rigidities. The market for both products and factors are thus assumed to clear by price adjustment in each period – where however the "period" may be interpreted as extending over several years.

The ELIAS model, as well as separate submodels developed by Bergman for different sectors of energy use, has been extensively used for the analysis and evaluation of Swedish energy policy. Besides the studies resulting from the common project already mentioned above (cf. Bergman-Mäler 1981, 1983) the model was used for evaluating the probable impact of discontinuing nuclear power production in Sweden (Bergman 1981).

The second model, ISAC, is a 36 sector disequilibrium model, developed by Ysander, Nordström and Jansson at the Industrial Institute for Economic and Social Research (IUI) in Stockholm, building on an earlier static model at the institute resembling the above mentioned Treasury model. By incorporating va-

[2] For a discussion of other models, mainly developed for pedagogical purposes, cf. Lybeck et al. (1984).
[3] *E*nergy, *L*abor, *I*nvestment *A*llocation and *S*ubstitution.
[4] *I*ndustrial *S*tructure *A*nd *C*apital Growth.

rious kinds of rigidities in price and wage formation it allows for disequilibria both in foreign trade and in the factor markets. The behavior of local governments, who employ about a quarter of total labor in Sweden, is endogenously determined by way of an integrated submodel. A fairly detailed treatment of central government taxes and transfers makes it possible also to use the model for simulating various kinds of fiscal policies and stabilization regimes.

The ISAC model has been used for several different kinds of policy evaluation studies. The studies concerning energy policy have already been mentioned (cf. Ysander 1983a, Nordström-Ysander 1983 and Ysander-Nordström 1983). A medium-term macroeconomic forecast was carried out by model simulations in 1979 (Ysander-Jansson-Nordström 1979). The long-term interrelation between industrial structural change and government expansion was analyzed in a study in 1980 (Nordström-Ysander 1980). Finally the efficiency of fiscal policy and wage policy in controlling local governments has been analyzed by model simulations (Ysander-Nordström 1985).

The Swedish Treasury has also recently raised its ambitions in regard to medium-term models. Partly based on the experience of the ISAC model and with the aid of two of its authors, an aggregate dynamic model, incorporating for the first time price and wage formation, was developed within the Treasury during 1982 and used for the 1984 medium-term survey. (Cf. Nordström 1982, Ministry of Finance 1984).

2. International Trade and Structural Change

A common starting point for the modelling work is the fact that Sweden is a small open economy which in both models is interpreted to mean i.a. that the supply of imports is perfectly elastic and that developments in the world economy can be treated as exogenously given.

A major task for the macromodels is to represent as accurately as possible the mechanisms by which price and demand changes in the world market are transmitted to the Swedish economy and there induce structural adjustment.

The transmission is modelled by export- and import-functions for the tradable goods sector of Swedish industry. In both models the level of export and import for each type of commodity depends on relative price and on the demand in world market and domestic market, respectively. Econometric estimates of price and demand elasticities are used, directly in ISAC and modified by theoretical considerations in ELIAS.

The prices of Swedish producers can thus deviate from those of foreign competitors. In Bergman's competitive equilibrium model this is explained by the assumption of "country-specific" commodities (cf. Armington 1969), while ISAC assumes price-setting producers in monopolistic competition adjusting their profit margin with regard both to the competitors' prices and to capacity utilization.

This specification of export and import functions must, however, be regarded a very rough approximations. There are reasons to expect that in some cases it is th relative change rather than the level that will be a function of relative price development. The elasticity can moreover be expected to vary depending both on specific competitive conditions limiting the range of price competition and on general business conditions. The market reactions to price increases and price decreases are not necessarily symmetrical in amplitude and time-structure. A particular form of specification bias is also implied by the omission of all transaction and marketing costs other than price.

In the macromodels these behavior patterns are aggregated over different commodities and different time-phases which means i.a. that sight is lost of the changes in commodity composition within each aggregate. This may in itself make it difficult to interpret the price elasticities. We may be comparing aggregates of different compositions – e.g. domestic sales versus imports of a certain composite good or exports versus world market trade – which means that the measured changes in relative price may in part not be due to different price development for individual goods but simply reflect the different composition of the aggregates. Changes in composition may moreover call forth changes in the aggregate rate of price change even without a changing rate in any individual line of goods. There is e.g. some evidence indicating that Swedish firms react to a profit squeeze by concentrating their marketing efforts on those less contested markets and commodity lines where they can raise prices without losing too many customers. The fact that the estimated relative price elasticities are rather low in magnitude – between 1 and 2 – may therefore partly be explained by marketing efforts being concentrated to price-insensitive subgroups within each aggregate.

The price elasticities used may thus be biased particularly in measuring effects of general relative price changes. However, sensitivity analysis seems to indicate that for a reasonable range of elasticity values the demand growth will still exert a dominant influence (Nordström 1982).

Changes in world market growth and/or price can be expected to affect Swedish industry in several stages. The immediate impact on sales will first be reflected in changing factor earnings – wages and profits. It will be further buffered by a changed rate of scrapping of old production capacity, and in ISAC – with producers in monopolistic competition – also by modified price-setting in foreign and domestic markets, respectively. By changing price and profit expectations the transmitted signals will in the next stage also influence investment behavior and by that the long-term resource allocation and industrial structure.

3. The Dynamics of Capital Accumulation

The time-lags and inertia in structural adjustment are in both models represented by distinguishing the various vintages of productive capital. Capital is thus modelled as being "putty-clay". In the ex ante production functions there are no restrictions on input substitution. The shares for capital, labor and energy can be freely selected in order to minimize cost in terms of expected input prices. Once installed however the capacity becomes "clay", i.e. the capital ratio is fixed and no further substitutions between major input aggregates are possible. At any given point of time the total production capacity of an industrial sector will thus be made up of a long line of different vintages, whose diverse input ratios "embody" the price expectations current at the time each vintage capital was invested. As a consequence the ex post functions will exhibit decreasing returns to scale in labor even when the ex ante production functions are linearly homogeneous.

The vintage structure of capital explains both why structural adjustment to world market changes is slow and why complete specialization is avoided even in competitive equilibrium models like ELIAS.

A long-term deterioration in the international competitive position of a particular branch of Swedish industry will only gradually be reflected in a shrinking of the productive capacity of that industry. The structural adjustment will come about by an increase in the scrapping of old plants and a decrease in the investment in new capacity. The funds and real resources thus released will successively be transferred to other and move profitable lines of production. In the same manner increased capital costs, caused e.g. by raised real interest rates, will only gradually become fully reflected in changed capital ratios.

The structural adjustment will in general proceed somewhat slower in ISAC than in the competitive equilibrium world of ELIAS, since "sticky" pricing and an evening out of utilization between plants – due to technological reasons and employment considerations – will delay the weeding out of plants earning a zero or negative quasi-rent. On the other hand, the decline in sales and profit will for the same reason be correspondingly greater and will react both on prices and investment and thus reinforce adjustment. Altogether adjustments in ISAC will, however, be somewhat slower and carried out on a lower profit level than what would be the case in a perfectly competitive world.

Since both models assume a fixed exchange rate regime, a deterioration of the terms-of-trade by e.g. import price increases not matched by price increases in the export markets, will not be accommodated by a changing exchange rate. Adjustment to the diminished real national income must instead come about through lowered real wages and various kinds of import substitutions. Tendencies of rising deficits in the balance of payments will in the ELIAS model be automatically countered by a rise in domestic saving, while in ISAC this would normally require government intervention. In both models the change in import prices would modify the pattern of input prices and thus influence both the choice of technology and the allocation of investment resources. Changed profit expectations and capacity utilization will also affect investment decisions in ISAC, where these decisions are represented by investment functions, while in ELIAS the total gross saving ratio is exogenously given.

4. The Energy Economy

The experience of the oil price hikes in the 70s made it particularly interesting to model and measure the effects on the Swedish economy of changes in the international energy markets. As already mentioned the concern about the impact of oil price changes was in fact one of the major reasons for starting the modelling work.

A principal ambition in both models has therefore been to represent as accurately as possible the structure of energy demand both by households and by industry. In modelling industrial energy demand technological choices are represented in a way that can be interpreted as reflecting a multistage decision-making. In ELIAS nested CES-functions are used. The first choice is between a capital-labor composite and an energy composite. In the second stage the substitution possibilities within the composite commodities are exploited, capital versus labor and electricity versus fuels, respectively. Once these ex ante choices are made, the input shares are regarded as fixed for that vintage.

In ISAC an ex ante choice is made regarding the input shares for capital, labor, electricity, fuels and other intermediate goods. Within the fuels aggregate, however, ex post substitution can be made between oil, coal and domestic fuels. The rationale behind this assumption is that the choice of fuel often has only limited effects on the rest of the installed production technology.

This relatively detailed specification of industrial energy demand combined with the vintage capital approach makes it possible to study the technological adjustment necessitated by e.g. sudden oil price increases. One of the main results of the energy policy studies mentioned earlier was to show that a major part of the macroeconomic impact of oil price hikes on the Swedish economy is due to various kinds of economic inflexibility. The ELIAS simulations were particularly concerned with the technological inflexibilities and with ways of designing and investing more flexibility in regard to energy use into the economy. In the study employing the ISAC model which incorporates a good deal of price and market inflexibilities as well as a wide range of government policy instruments, interest was more focused on designing and safeguarding flexible policies to compensate for the rigidities of the markets (Ysander 1983b,c).

5. Classical versus Keynesian Modelling

We have so far mainly dealt with the aims and features common to the two models – the way price signals and shifts in the world markets are transmitted through foreign trade into Sweden's small open economy and there gradually transform capital structure, production technology and energy use through the successive replacement of old vintages.

Of at least equal importance and interest, however, are the conceptual differences in approach between the models. They, in fact, seem well suited to exemplify the methodological dissimilarities between the two traditions in macromodelling often, somewhat inaccurately and simplistically, identified as "classical" and "Keynesian".[5]

[5] For a stringent but more narrow definition of classical and Keynesian macromodels, cf. e.g. Sargent (1979).

The label "classical" is here meant to refer not so much to the common foundations of neoclassical analysis as to some characteristic premises and priorities shared both by "old" and "new" classical economists. Models in this tradition are usually built on the assumption of a competitive equilibrium with immediate market clearing, and priority treatment is often given to the supply side of the economy, with demand, and particularly public demand and policy, dealt with in a more summary fashion. In the unavoidable trade-off between on the one hand theoretical consistency and coherence and on the other hand empirical verification and realism, model builders in this tradition tend moreover to favor the former.

The "Keynesian" modellers usually lean the other way and therefore tend to incorporate various forms of marked rigidities and limits to competition into their macromodels, even at the risk of having to accept "adhocery" and divergencies from the postulates of rational behavior. The recognition of market failures is moreover seen as implying the need for government intervention and demand management, requiring a more detailed specification of demand structure and policy instruments.

In terms of this oversimplified dichotomy ELIAS should undoubtedly be called a classical model while ISAC seems well qualified for the Keynesian label.

ELIAS is a competitive equilibrium model, assuming immediate market clearing. It extends the earlier analysis of multisectoral growth by Leif Johansen (1960) mainly by "opening" the model and explicitly formulating the linkage to the world markets and by taking account of the dynamics of capital accumulation by using a vintage representation of technological choice. The emphasis is on analyzing supply side developments although household demand is represented by a linear expenditure system, common with ISAC. Government demand is taken as an exogenously given aggregate and no taxes or any other policy instruments are explicitly specified. The specification of the model is to a large extent directly derived from neoclassical economic theory. To achieve this some empirical accuracy must be sacrificed by substituting calibrations and "guesstimates" for econometric estimates. It is also in line with the ambition of theoretical transparency to keep the model relatively small and compact. The specification has moreover been formulated so as to make it possible to use rational expectations and some simulation experiments have indeed been based on this assumption.

In ISAC, belonging to the tradition of Keynesian disequilibrium models (cf. Barker 1976), the modellers have gone far in the opposite direction, trying to incorporate various kinds of adjustment obstacles – besides immalleable vintage capital also sticky wages and prices, cash-flow restrictions on investment financing and the inertia and lags observed in both private and local government consumption. Government demand and policy instruments are specified in some detail with local government budgets being endogenously determined. The model is "extra-Keynesian" in the sense that it has been designed to make it possible to take into consideration also rigidities in local governments and various possible constraints on central government policy. These ambitions together with the desire to be able to exploit available statistics and survey data on individual industrial branches have made the model big and rather complex. This unavoidably entails the risk for inconsistencies and difficulties in tracing unambiguously the effects of variations in the exogenous variables. As far as possible the implementation of the model has been based on econometrical estimates.

These differences between the models will affect the analysis of structural change in important ways. With the equilibrium assumptions of the ELIAS model we can study in isolation the long-run impact of world market disturbances on Sweden's industrial structure with the immalleability of capital as the only factor of inertia and without getting the causal picture blurred and distorted by central government interventions and endogenous local government reactions. The picture will be clear but partial in the sense of neglecting interrelations between private and public sectors and focusing only on the industrial structure.

With the ISAC approach market disequilibria – surpluses or deficits of foreign exchange, public budgets, production capacity and labor – will be a normal feature of model projections and will have feed-back effects in the form of price modifications, rationing of supply and changing patterns of local government allocations. From a disequilibrium situation the model economy may finally – with the help of or despite economic policy measures – fetch up in a new equilibrium or steady-state growth, barring new disturbances. The adjustment path will in most cases affect the final structure in several important ways. To study this interdependence between short-run instability and long-term structural growth is indeed one of the main purposes of the ISAC model.

6. Market Behavior

The differences between the two models are best exemplified by looking at the functioning of the various markets.

In the *markets for tradable goods* flexible prices will guarantee equilibrium between demand and supply in ELIAS. Total demand is composed of export demand, investment demand, household demand and intermediate demand. The intermediate demand – apart from energy – is determined by input coefficients, which are the same for all vintages. The demand for investment goods is derived from the investment within the different sectors. Household demand is determined by the aggregate consumption propensity – assumed constant in ISAC and indirectly determined in ELIAS by the given gross saving ratio. The distribution of total household demand between different commodities is in both models determined by a linear expenditure system. Supply by the profit maximizing and price-taking firm will in the assumed competitive equilibrium be set at levels equating market price with marginal cost.

In ISAC the producers are instead price-setters in monopolistic competition. Their profit-maximization is moreover constrained by i.a. employment considerations, forcing them to try to keep an even rate of utilization between the plants in operation,[6] although successively scrapping units with a negative quasi-rent. The

[6] For a discussion of the modelling of constrained optimization in disequilibrium model, cf. i.a. Bureau-Miqueu-Norotte (1984).

producers may also discriminate in their price-setting between foreign sales and the home market. In both cases their pricing can be said to reflect an attempt to cover costs, computed as average variable cost plus planned depreciation and a target rate of return on installed capital. This "cost-price" is then modified to take account both of foreign competition and of variations in capacity utilization. As already exemplified above, the "sticky" prices and the evening out of capacity use will tend to slow down structural adjustment, while the possibility of price discrimination will sometimes mean that home consumers subsidize foreign sales by inflated domestic prices. Production is assumed to adjust to actual sales with proportionate variations in stock-keeping. Disequilibria will thus occur in the form of over- and underutilization of capital. A slow adjustment of capacity will take place, since investments are determined by the rate of utilization as well as by past profit performance.

Supply in the *labor market* is exogenously given in both models. Demand is derived from the levels of current production in the private and public sectors.

The mechanism for wage determination is, however, very different in the two models. In ELIAS the general wage level is determined as an equilibrium price, guaranteeing full employment. In order to take account of labor heterogeneity, however, an exogenous wage structure is imposed by multiplying the general wage with sector-specific coefficients.

Wage-setting in ISAC is instead modelled as the outcome of a negotiating process, where wage earners try to get compensated for both inflation and productivity gains but where the final result will be modified by current market conditions, i.e. unemployment. Long-run wage adjustment will thus be reinforced by the change in total employment, resulting from the current bargaining. For public employees a one-year lag in wage settlement, relative to the private sector, has been usual and has been assumed to continue in the future. Apart from this time-lag in compensation, there are no sector specific wage differentials in ISAC.

There are no explicit financial markets in either model. The rate of exchange is in both models assumed to be exogenously determined. The treatment of *capital transactions,* i.e. investments is however different.

In ELIAS the gross saving ratio in the economy is assumed to be given. Since the size of the balance of payment deficit or surplus is further assumed to be set by government at some target value, total domestic saving or investment will be determined along with domestic production. Investment resources are then allocated between the different production sectors in proportion to the expected relative profitability of new capacity within each sector. In principle the market rate of interest is determined as the "cut-off" rate which makes total investment demand equal to the given supply of investment resources. In practice, however, a slightly more roundabout way is employed to ensure also that rates of return will tend to equalize between sectors despite the absence of returns to scale in the ex ante production functions. The capital market implicit in this modelling can perhaps be described in terms of a given "loanable fund" with the market rate of interest determined by the "marginal efficiency of capital".

In ISAC the rate of interest is instead assumed to be determined by conditions in the international financial markets. The small open economy with its heavy dependence on international finance is supposedly unable under a fixed exchange rate regime to deviate substantially and persistently from international standards of yield. This assumption could be interpreted as implying a central bank policy of monetary accommodation aimed at keeping the public's demand for money satisfied at the given rate of interest. The internationally determined rate of return requirement is further transformed in the model to a particular rate for each branch of industry by taking account of differences in depreciation rate, tax treatment and solidity.

Investments in the different sectors are in ISAC functions of capacity utilization and past "excess" profits. These excess profits are measured relative to the user cost of capital, with the rate of return requirement as a major component. The required rate of return will influence the firm's pricing and investment and by that also its saving, but will not affect the saving ratio for households. The main burden of adjusting total domestic saving to avoid surpluses or deficits in external payments will thus fall on the public budgets, particularly the state budget.

7. Government Behavior and Stabilization Regimes

In deterministic equilibrium models, where instant market clearing is insured by flexible prices, stabilization and incomes policies are almost by definition superfluous. There is consequently no need for a detailed specification of public budgets in order to pinpoint the various available policy instruments or to study the possible destabilizing effects of endogenous local government action. It is thus consistent with the modelling ambition in ELIAS to treat public spending as an exogenous aggregate and to define taxing only implicitly by the assumed constant ratio of gross saving in the economy.

The opposite is true of the disequilibrium model, ISAC. The interrelation between stabilization problems and structural change is here in the focus of interest. In order to be able to study these interrelations a detailed account of public allocations, of taxes and transfers and a brief accounting of their distributive effects are built into the model. By the use of a submodel, LOGOS, (cf. Ysander 1985), the behavior of local governments is moreover endogenously determined.[7]

[7] The local government sector in Sweden is, by international standards, big, financially independent and expansive. It employs more people than the manufacturing industry and thus exerts a dominant influence on wages and labor market conditions. During the 70s the overall net effect of variations in local government activity appears to have been procyclical and destabilizing. The LOGOS model is a ten-equation system determining five kinds of service expenditures, two kinds of transfers, investments, borrowing and, as a residual, the local income tax rate (cf. Ysander (forthcoming)). Its incorporation in the macromodel has made it possible to explain i.a. the inefficiency of grants policy in changing the local government share of total consumption and the possibility that local government may by itself generate cyclical movements in wages and prices and reinforce disturbances from abroad (Ysander-Nordström 1985).

One of the main purposes with the ISAC model has been to use it for policy evaluation. Apart from the studies of energy policies, already mentioned, the main emphasis has been on medium-term stabilization regimes, ranging from various fiscal policies and control measures directed at local governments to wage controls and other income policies. In Keynesian terms one could say that the exogenous interest rate in ISAC combined with inelastic household saving and sticky wages make fiscal policy needed. At the same time adaptive expectations, demand-sensitive investment functions and the absence of financial crowding-out usually ensure that fiscal measures will have a short-run impact on employment before this effect is overtaken by a wage-propelled inflation.[8]

8. Comparing Models

Having examined the various characteristics of the two models we will try to sum up the discussion by comparing the models in terms of realism, relevance, resilience and robustness.

Econometric methods are supposed to help us in determining to what extent a model explains real life events satisfactorily. Unfortunately we are seldom able – even when we try – to use econometric tests also to choose between alternative models – alternative simplifications of reality.

Some of the conceptual differences between ELIAS and ISAC belong to the core of current macroeconomic controversy. The attempts made so far to test econometrically key assumptions and parameters do not seem to have yielded very conclusive or generally convincing results. One cannot escape the impression that the macroeconomic argument is often not just about the right answers but in equal measure about the right choice of questions. We argue about *realism* even when we really disagree about *relevance*.

The different aims of our two models exemplify the range of relevant questions. The interest in ELIAS is mainly focused on analyzing as clearly as possible the mechanisms of long-term structural adjustment, while largely abstracting from the short-term aberrations and stabilization problems on the demand side. ISAC instead aims at providing quantitative answers to these short-term questions and investigates how the management of short-run problems can affect the process of long-term structural change. The choice of question may depend on your time perspective or special interests but can also be contingent on whether you think short-run problems matter and can be meaningfully treated by public policy.

[8] In practice the stabilization policy experiments are usually set up in the following manner. The multiplier effects of variation in policy instruments on certain chosen target indicators are measured for a predetermined medium-term period in relation to a reference growth path. The results are then tested to see if the dynamic structure can be viewed as approximately linear in a local neighborhood of the reference path. If this is possible it is a straightforward matter to compute "optimal policy packages" or the trade-off between policy instruments in terms of the target indicators. (For a discussion of linearized model analysis and the application of control methods, cf. e.g. Kuh-Nease 1982 and Chow 1982).

The problem of *resilience* is important both in equilibrium and disequilibrium models. With resilience we here mean the ability of an economic system to absorb temporary outside shocks and adjust back to a stable growth path close to the one initially given without intervention in the form of policy changes. Intuitively one tends to presume that models with a Walrasian market equilibrium like ELIAS will have a high degree of resilience. The flexible prices will buffer the shocks and there will be no risks that the feed-back of disequilibria will further reinforce the deviation from the original growth path. In the case of ELIAS the intuition seems true as far as can be judged from the simulations carried out. For ISAC it is more difficult to know what to expect. Experiments show, however, that feed-back mechanisms in the labor and capital markets will tend in time to restore full employment of both people and capacity. There is no corresponding mechanism though for the balance of payment.

Another important aspect of models concerns *robustness,* i.e. their structural stability as related to the parameters of the models. To what extent are the constant parameters really constant and, if not, how sensitive is the model performance to parameter changes?

A number of sensitivity tests for parameter changes has been carried out for ISAC. They show i.a. that even though the model performance changes continuously and not very dramatically with variations in single key parameters, quite large deviations may occur if several parameters are allowed to vary simultaneously. The problem with these kinds of test is of course that we have no reliable way of deciding in advance which are the parameters most important to the dynamics of the system. Methods have been developed for measuring the relative importance of parameters in terms of a linear approximation of non-linear models (cf. e.g. Kuh-Nease 1982). These methods however still leave us with the question how valid the approximation is outside the chosen reference path.

Another aspect of the problem concerns the reliability of econometric estimates, when applied deterministically in new economic situations (cf. Lucas 1976). How far do we go wrong by neglecting not only random variations but also adaptive changes by the economic agents reacting to new conditions? These are obviously important questions, particularly for models like ISAC, which depend to a great extent on econometric estimates. The fact that both external conditions and economic policy in Sweden have changed drastically since the period of estimation increases the uneasiness. We have so far no way of resolving this dilemma. We can only hope that errors of this kind will not be able to change the sign and direction of the results. A reasonable conjecture could perhaps be that the implications are less important for short-run behavior than for long-run performance (cf. Brandsma-Hughes Hallet 1984). Whatever the truth of this much work remains before we can be sure of having achieved reasonably reliable and robust models.

References

Ancot, J. P. (ed.), 1984, *Analysing the Structure of Econometric Models,* Martinus Nijhoff Publishers, the Hague.

Armington, P. S., 1969, "A Theory of Demand for Products Distinguished by Place of Production", *IMF Staff Papers.*

Barker, T.-S. (ed.), 1976, *Economic Structure and Policy,* London.

Bergman, L., 1981, "The Impact of Nuclear Power Discontinuation in Sweden", *Regional Science and Urban Economics.* Vol. 11.

Bergman, L., 1982, "A System of Computable General Equilibrium Models for a Small Open Economy", *Mathematical Modelling,* Vol. 3.

Bergman, L. and Mäler, K.-G., 1981, "The Efficiency - Flexibility Trade - Off and the Cost of Unexpected Oil Price Increases", *Scandinavian Journal of Economics,* Vol. 83, No. 2.

Bergman, L. and Mäler, K.-G., 1983, "Oil Price Uncertainty and National Energy Strategies"; in Ysander (1983c).

Brandsma, A. S. and Hughes Hallet, A. J., 1984, "The Structure of Rational Expectations Behaviour in Economics: An Empirical View"; in Ancot (1984).

Bureau, D., Miqueu, D. and Norotte, M., 1982, "The Disequilibrium Theory in Macroeconomic Models: a small scale model"; in Ancot (1984).

Chow, G. C., 1982, "Analysing Econometric Models by Control Methods"; in Chow - Corsi (1982).

Chow, G. C. and Corsi, P., 1982, *Evaluating the Reliability of Macroeconomic Models,* John Wiley & Sons.

Gramlich, E. M. and Ysander, B.-C. (eds.), 1985, *Control of Local Government,* IUI, Stockholm.

Jansson, L., 1983, "A Vintage Model for the Swedish Iron and Steel Industry"; in Ysander (1983b).

Johansen, L., 1960, *A Multisectoral Model of Economic Growth,* North Holland, Amsterdam.

Klein, L. R., 1982, "Economic Theoretic Restrictions; in Chow - Corsi (1982).

Kuh, E. and Nease, J., 1982, "Econometric Model Diagnostics"; in Chow - Corsi (1982).

Le Van, C., 1984, "Specification of the Import Function and the Determination of Short and Long Term Equilibria in Macroeconomic Models"; in Ancot (1984).

Lucas, R. E., 1976, "Economic Policy Evaluation: A Critique", *Supplement to Journal of Monetary Economics,* April.

Lybeck, A. J. et al., 1984, "A Comparison of the Dynamic Properties of Five Nordic Macroeconometric Models", *Scandinavian Journal of Economics,* Vol. 86, No. 1.

Ministry of Finance, 1976, *Medium Term Planning in Sweden - The System of Models,* Stockholm.

Ministry of Finance, 1984, *The 1984 Medium Term Survey of the Swedish Economy - The System of Models,* Stockholm.

Nordström, T., 1982, *ISAC: Modellbeskrivning och kalkylexempel 1981–1985* (Model description and simulation exercise 1981–1985), Ministry of Economics, Ds E 1982:2.

Nordström, T. and Ysander, B.-C., 1980, *Offentlig service och industriell tillväxt. Perspektivskisser av svensk ekonomisk utveckling 1950–2000* (Public service and industrial growth - Alternative paths of development for the Swedish economy 1950–2000), IUI Research Report No. 11, Stockholm.

Nordström, T. and Ysander, B.-C., 1983, "Oil Prices and Economic Stability - Simulation Experiments with a Macroeconomic Model"; in Ysander (1983c).

Richard, J.-F., 1982, "Exogeneity, Causality and Structural Invariance in Econometric Modelling"; in Chow - Corsi (1982).

Sargent, T. J., 1979, *Macroeconomic Theory,* Academic Press, New York.

Ysander, B.-C., 1983a, "Oil Prices and Economic Stability: The Macro-Economic Impact of Oil Price Shocks on the Swedish Economy", *Energy,* Vol. 8, No. 5.

Ysander, B.-C. (ed.), 1983b, *Energy in Swedish Manufacturing,* IUI, Stockholm.

Ysander, B.-C. (ed.), 1983c, *Energy and Economic Adjustment,* IUI, Stockholm.

Ysander, B.-C., forthcoming, *A Model of Local Government Behavior* IUI, Stockholm.

Ysander, B.-C., Jansson, L. and Nordström, T., 1979, "Utvecklingsvägar för svensk ekonomi 1978–1985" (Economic choices for the 80s); in *Kalkyler för 80-talet* (Forecasts for the 80s), IUI, Stockholm.

Ysander, B.-C. and Nordström, T., 1983, "Energy in Swedish Manufacturing 1980–2000"; in Ysander (1983b).

Ysander, B.-C. and Nordström, T., 1985, "Local Authorities, Economic Stability and the Efficiency of Fiscal Policy"; in Gramlich-Ysander (1985).

Åberg, C.-J., 1971, *Plan och prognos* (Planning and Forecasting), SOU 1971:70, Government Printing Office, Stockholm.

CHAPTER V

Measuring Energy Substitution An Introduction

For a small open country like Sweden the ability of the manufacturing industry to adjust rapidly and smoothly to changes in relative world prices is of crucial importance. The oil price hikes in the 70s tested this ability in a dramatic fashion, creating at the same time a particularly good opportunity for studying the mechanisms and the adjustment problems involved in industrial factor substitution. Economists all over the world have hastened to exploit this opportunity and, as a result, our knowledge of industrial production structure and factor adjustment has increased considerably over the last few years.

The papers assembled in this volume, focusing on energy use in Swedish manufacturing, all share this common aim of mapping and measuring industrial adjustment to price changes. However, as appropriate for a still developing research area, they try alternative approaches, using different models and analytical techniques.

Mechanization and energy use - the postwar experience

From the end of the war and up to the first oil price hike the ongoing mechanization had a dominant influence on energy use in Swedish manufacturing. This is one of the main lessons to be learned from the first paper — by Joyce Dargay — tracing energy prices and energy use in the postwar period.

* The Industrial Institute for Economic and Social Research (IUI), Stockholm.

Mechanization meant continuous substitution of both capital and energy for labor in industrial production and it usually also meant electrification. The specific use of electricity (i.e., electricity input per unit of output) in manufacturing thus increased steadily during the 50s in spite of sharply rising electricity prices, stagnated in the 60s when prices were falling both in absolute terms and relative to other energy prices, but started mounting again in the latter part of the 70s, when electricity prices were catching up with other energy prices.

The other dominant postwar trend in energy use, viz. the switch from solid fuels to oil, can also partly be interpreted as a way of saving labor, whose wages, up to 1973, grew steadily faster than both capital and energy costs. Although the coal price paid by manufacturing industry tended to keep pace with the oil price, although with a certain time-lag, the labor costs involved in handling coal made oil advantageous to use, at least up to the first oil price hike. In terms of specific usage the main switch to oil occurred already in the early 50s. Although prices for heavy oil were completely stagnant up to the middle 60s, the specific oil use in manufacturing did not increase further during this period. The first oil price hike led to a considerable drop in specific oil usage, which then further declined at the end of the decade.

The continued rapid decrease in the use of solid fuels together with the curtailed oil use during the 70s, resulted in a steady reduction in total specific energy use from the middle of the 50s. About a quarter of this total energy saving was due to the changing branch structure. With a few exceptions — printing, chemicals and shipbuilding — the energy/output ratio fell in all manufacturing branches.

The dominant influence of mechanization and labor saving on industrial energy demand underlines the importance of analyzing energy use within the framework of production models, incorporating all the substitution possibilities between different inputs.

There are wellknown reasons for distinguishing also between manufacturing branches with different technologies, different capital/ output and energy/output ratios. The price data collected by Dargay indicate another and less discussed reason for disaggregation. They show i.a. that in 1968 the energy intensive branches — pulp and paper, chemicals, primary metals and non-metallic mineral products — paid on the average 1/7 less for heavy oil and only half for electricity compared to other branches. After 1973 there is, however, a considerable convergence of energy prices between branches — in the case of petroleum products partly due no doubt to a shortening of contract lengths and a reduced significance of rebates to large consumers. The energy intensive branches thus faced a rise in energy prices during the first half of the 70s that was on the average 40 % larger than that experienced by the other sectors.

Alternative ways of measuring factor substitution

Any attempt to explain and measure the possibilities of factor substitution in industry must at the start make two kinds of basic choices: a choice of aggregation level and a choice of adjustment paradigm.

The choice of aggregation level involves at least three different dimensions of the production structure: technologies, factors and firms.

There is obviously a limit to the degree of technological detail that could and should be included in an economic production model. The use of approximate (or "generalized") descriptions of technologies in the form of production functions and of aggregations of those over different technologies has the advantage of small data requirements and computational ease. The disadvantage is the introduction of approximation and aggregation errors in the numerical results. The risk of distorted and biased results caused by the functional approximation is particularly great if the functions used are such as to place a priori restrictions on the factor substitution to be measured. Fortunately, theoretical research in the 60s and 70s has provided us with a family of flexible functional forms — like the generalized Leontief and the translog — which i.a. do not assume any restrictions on substitution elasticities. (For a short survey of these developments cf Field-Berndt, 1981). Simulation experiences indicate, however, that the aggregation error may still be large enough to make it difficult to get stable and consistent estimates on substitution relationships from aggregate descriptions of technologies (cf. Kopp-Smith, 1981).

The conditions for subsuming different machines and constructions under an aggregate capital measure in the production function, or for aggregating different labor inputs, have been thoroughly discussed over the past years. In the present studies we have to deal with yet another kind of factor aggregation — the aggregate treatment of different sources of energy: petroleum products, solid fuels and electricity. This brings into focus two new and interesting questions. Is the supply of energy to industrial plants so flexibly designed that energy production from different sources, the primary energy allocation system, can be treated as completely independent — or "separable" — from other technological decisions?

The second question is concerned with causation in the opposite direction. How much will the optimal internal allocation of "composite capital" depend on the relative price of energy and the way it is produced? As stated above already, our intuitive reading of the postwar experience indicates a rather strong dependence both ways. The rapid substitution of oil for coal in the 50s was probably not only motivated by labor saving but in part dictated by the new technologies imported from the U.S. On the other hand it seems evident that the oil price hikes in the 70s had an important and different impact on the profitability of different types and vintages of existing production capital. Any results of studies dealing with aggregate capital and energy in the conventional way must therefore be interpreted with great caution.

Aggregating over firms adds yet another problem dimension, since i.a. the rate of utilization in the firms may vary with their different positions and strategies in the market and the resulting allocation of production between firms may not be stable over time.

Having decided on the proper level of aggregation, one is still left with the choice of adjustment paradigm, i.e., the decision about how to model the way production is adjusted to market changes.

One part of this choice is concerned with modeling the market on which the producers are supposed to operate. If one should take into account the particular kind of say oligopolistic market structure involved, should discern both sides of the mutual adjustment of supply and demand, and consider also the possibility of disequilibrium pricing, the modeling ambitions could easily outrun the available resources for estimation. The market is therefore usually treated in a very simplified manner, e.g., by assuming the producers to be price takers and/or to operate within fixed market shares.

Changes in market conditions should call forth two kinds of supply adjustment. The short-run adjustment is concerned with accommodating the market changes within existing capacity by changing the current production and with that the cost-minimizing input demands. The long-run adjustment is initiated by capacity changes, due both to technological changes and to investment/ scrapping activities. Since adjustment is costly and time-consuming the relevant decisions will stretch far into the future, and will depend on expectations which in turn may be built on historical experience going far back in time. Because of the obvious difficulties involved, very few attempts have been made so far to model this adjustment process explicitly as an intertemporal optimization under uncertainty. In most studies the dynamic element is simply represented by some rather ad hoc lag structure or accelerator relation. (For a thorough discussion of the various stages of dynamic adjustment representation, see Berndt, Morrison, Watkins, 1981). Indeed the majority of studies on factor substitution documented so far are based on static models which completely disregard the existence of adjustment constraints, assuming instead full and instant adjustment. As for technical change, whether embodied or disembodied, it is usually, for the sake of computational convenience, treated as neutral. This means disregarding the possibility — observed in many econometric studies — that the rate and direction of productivity change may depend on relative factor prices.

Most of these problems of simplification and model choice are exemplified in the four studies of industrial substitution possibilities contained in Part II. Table 1 gives some indication of the variety of models and methods used by the different authors. The first five columns are concerned with characteristics related to aggregation, the next two columns reflect the choices of "adjustment paradigm" while the last indicates the method of estimation.

Dargay uses time-series data to estimate factor cost shares for twelve manufacturing branches. Her translog cost functions include capital, labor, intermediate goods and energy as arguments, with the aggregate energy input being alternatively measured directly in terms of physical energy units or estimated indirectly as a cost-minimizing mix of primary energy inputs. Her model is essentially static with Hicks neutral technical change. Both a homothetic and a non-homothetic functional form were tried, with the non-homothetic formulation giving more significant and consistent results. Dargay did not, however, succeed in producing separate estimates of rates of return to scale and technical change. An FIML estimation program was used in two stages — firstly to estimate the cost-minimizing energy mix and secondly to estimate the cost-shares of the aggregate factors.

The Jansson study of the Swedish iron and steel industry is based on the same time-series data as the Dargay study, deals with the same aggregate factors — although assuming proportionality in

Table 1 Four approaches to measuring factor substitution

Author	Data	Coverage	Level of aggregation	Production factors in estimation	Form of cost function	Adjustment constraints	Technical change	Method of estimation
J. Dargay	Time-series 1952-76	Total manu-facturing	Twelve branches	C,L,M,E, E(e,o,s)[a]	Translog	-	Hicks neutral	Two-stage FIML
L. Jansson	Time-series 1952-75	Iron and steel	Capital vintages of branch	C,L,E[a]	-"-	Vintage capital, demand and profit de-velopment	-"-	FIML
L. Hultkrantz	Cross-section statistics and engineering data, 1979	Wood, pulp and paper in northern Sweden	Production activities within plants	I,L,M,e,o[a]	Linear	Supply constraints, current capacity and available investment options	-	LP
S. Lundgren	Engineering data	Iron and steel	Production activities	C,L,M, e,o,s[a]	-"-	Short run: capacity constraints Long run:-	-	LP

[a] C = Capital E = Energy I = Capital investment o = oil
L = Labor M = Intermediate goods e = electricity s = solid fuels

the use of intermediate goods — and used the same FIML estimation program. However, there are several important differences between his study and that of Dargay. In the Jansson study the different annual vintages of production capital are distinguished in the model. Even more important, Jansson's model, which is of the "putty-clay" type, attempts to explain adjustment in terms of gross investment and scrapping of production capacity, with the technology of the new capacity reflecting current factor prices. In such a dynamic model, "substitution effects" in terms of technological adjustment in new capacity to short-term changes in factor prices, may be overlaid and dominated by "vintage effects", resulting from adding new capacity to an existing stock which reflects techniques and prices over the past thirty years. Two inputs — like capital and energy — may then be substitutes in the technological sense and yet be complementary over time, even without non neutral technical change. In this way, the Jansson study reconciles the diverging and controversial results obtained in earlier studies of the elasticity of substitution between capital and energy in various manufacturing branches.

The two following studies both use a radically different approach. Instead of time-series data they use cross-section and/or engineering data and are then able to model individual production activities within the branch in question. For the same reason there is no need for them to try to aggregate the diverse kinds of primary energy resources. To be able to handle this mass of technological information, they are forced to linearize all relations, so that optimal production plans can be computed with linear programming techniques. While in the preceding studies elasticities of substitution could be computed from parameters of the estimated functions, they can now only be very roughly approximated by comparing the outcome of different runs of the LP-models.

Hultkrantz' study of the wood, pulp and paper industry in northern Sweden encompasses two periods and includes different packages of investment options for the two time horizons. The options are those currently considered by the firm at the time of the enquiry (1979). In terms of this multiperiod model Hultkrantz can define a concrete and specific meaning and measure for the distinction between short-run and long-run adjustment. A special feature of the Hultkrantz model is the fact that the paper and pulp industry is here embedded within a larger model, which takes explicit account of alternative uses of wood — for the sawing industry and more particularly for heat generation. One of his main conclusions, of great importance and relevance for current Swedish energy policy, is that only very drastic further increases in the relative oil price could make wood-based heating stations a serious competitive threat for the forest-products industries. This and related results are derived by maximizing the quasi-rents to industrial capacity and the price of stumpage subject to the constraints set by industrial capacity, investment opportunities and available volumes of wood of different kinds.

Lundgren's study of the iron and steel industry is entirely based on engineering data and blueprints for future technologies. His model is essentially a static one-period model with explicit capacity constraints. Long-run adjustment can be defined and measured by eliminating all capacity constraints. While Hultkrantz' experiments are based on maximizing profits or quasi-rents, Lundgren's simulations all deal with cost-minimization, holding the output mix constant.

Some numerical results

Four different ways to model reality lead to four different modes of designing questions about factor substitution — and imply four different types of answers. We will make no attempt here to survey or summarize the numerical results recorded in the four studies in Part II. The examples presented in Table 2 below merely serve the purpose of illustrating the variety of numerical experiments performed and of substitution mechanisms investigated.

Table 2 Elasticities of substitution - some numerical results*

Author	Type of elasticity	Branch	Energy(oil) - - Capital	Energy(oil) - - Labor	Oil - Electricity	Energy (oil)
J. Dargay	Allen partial elasticity of substitution = σ	Total manufacturing	$\sigma_{EC} = -1.43$[1]	$\sigma_{EL} = 0.12$[1]	$\sigma_{oe} = 0.21$[2]	$\varepsilon_{oo} = -0.29$[3]
	Price elasticity = ε	Wood, pulp and paper	$\sigma_{EC} = -0.59$[1]	$\sigma_{EL} = 0.02$[1]	$\sigma_{oe} = 0.22$[2]	$\varepsilon_{oo} = -0.28$[3]
		Iron and steel (Primary metals)	$\sigma_{EC} = -0.66$[1]	$\sigma_{EL} = -0.61$[1]	$\sigma_{oe} = 0.24$[2]	$\varepsilon_{oo} = -0.26$[3]
L. Jansson	-"-	Iron and steel	$\sigma_{EC} = 0.82$[4]	$\sigma_{EL} = 2.63$[4]		$\varepsilon_{EE} = -0.98$[4]
L. Hultkrantz	Arc cross-price elasticity = e (profit maximization under supply and capacity constraints)	Wood, pulp and paper in northern Sweden	$e_{Io} = -0.57$[5]	$e_{Lo} = -0.29$[5]	$e_{eo} = -0.72$[5]	$e_{oo} = -0.49$[5]
S. Lundgren	Arc cross-price elasticity = e (cost minimization for given output mix)	Iron and steel			$e_{eo} = -0.26$[6]	$e_{oo} = -4.3$[6]

1 Homothetic cost function, direct estimates.

2 Partial substitution effects, total energy consumption constant.

3 Total own-price elasticity, non-homothetic total cost function.

4 Elasticities of the ex ante production function

5 50% oil price increase, long-run adjustment including output change.

6 50% oil price increase, long-run adjustment without investment constraints, output constant.

The different approaches are reflected in different notions and measures of the elasticity of substitution.

The concept of elasticity of substitution, as originally introduced by Joan Robinson (Robinson, 1933), is intended to measure the ease of substituting between two inputs, when output is held constant. It is usually defined as the derivative of the ratio of two input levels with respect to the ratio of the two corresponding input prices. For a production function with two inputs, $Q = f(x_1, x_2)$, and the corresponding input prices, p_1, p_2, the elasticity of substitution can be written as:

$$\sigma_{12} = \sigma_{21} = \frac{d \ln(x_1/x_2)}{d \ln(p_2/p_1)}$$

σ_{12} here grows larger as substitution becomes easier. Also, when $\sigma_{12} > 11$, the cost share of input 1 becomes larger relative to the cost share of input 2 when input 2 becomes relatively more expensive.

With more than two inputs involved, however, different definitions of elasticity result from different choices of the ceteris paribus conditions under which the partial derivatives are obtained.

The most commonly used definition — the Allen-Uzawa partial elasticity of substitution — is simply a price cross elasticity weighted by the inverted value of the corresponding cost share:

$$\sigma_{ij} = \left.\frac{d \ln(x_i/x_j)}{d \ln(p_j/p_i)}\right|_{Q=\bar{Q}} = \frac{1}{p_j x_j / \sum_i p_i x_i} \cdot \left.\frac{d \ln x_i}{d \ln p_j}\right|_{Q=\bar{Q}} = \frac{1}{k_j} \varepsilon_{ij},$$

where k_j denotes the cost share of the jth input. In this definition all other inputs adjust optimally to the price change.

As shown in Table 2, Dargay's elasticity measures for total manufacturing show complementarity between capital and energy, while energy and labor appear to be relatively independent of each other. Oil comes out as a rather poor substitute for electricity, and also registers a low own-price elasticity.

Her results for the wood, pulp and paper industry, also shown in the table, are very similar to those for total manufacturing although the complementarity between capital and energy does not register as strongly for this branch.

For iron and steel and other primary metal industries the one main divergence in results, compared to the wood, pulp and paper industry, is the complementarity here registered also between energy and labor.

As for intermediate goods, Dargay's results seem to support the conclusion from Parks' earlier study of Swedish manufacturing 1870-1950, that capital and labor in most branches are not separable from intermediate goods (Parks, 1971).

In Jansson's production model for the iron and steel industry a distinction can be made between the "potential" substitution possibilities of the ex ante production function and the actual realized substitutions, which may to a large degree be determined by "vintage effects", i.e. by the inertia due to older capital vintages. This may explain why energy shows up in his study as a strong substitute in the more narrow technological sense for both capital and labor, while the opposite result is derived from Dargay's static model.

The elasticity measures recorded in the LP studies of Hultkrantz and Lundgren are quite different from those used in the preceding papers. Firstly, they are arc elasticities, which means that instead of being computed from parameters of estimated production functions, they are rough measures of average effects of intramarginal — and in fact quite drastic — price changes in the model simulations.

Secondly, there are in the simulations important constraints — concerning production capacity and raw material supply — on the adjustment of inputs. In this regard their elasticity measures are not so much related to the Allen-Uzawa elasticity as to the concept of "direct" elasticity of substitution, which holds constant other inputs than those directly concerned (McFadden, 1963).

Thirdly, what they compute are straightforward cross price elasticities and not elasticities of substitution, although these two concepts are closely related (cf. above).

Finally, in the case of Hultkrantz, the elasticities are not computed with output held constant, which means that the measured effects of input price increases are also influenced by shrinking total production.

This last point probably to a large extent explains why Hultkrantz finds energy to be a complement not only to capital but also to labor. For the case of regular neoclassic production functions it has been shown (Field-Allen, 1981) that a cross price elasticity with freely variable output can be defined as;

$$\eta_{ij} = \left.\frac{d \ln x_i}{d \ln p_j}\right|_{\substack{p_k = \bar{p}_k \\ k \neq j}} = \varepsilon_{ij} + k_j \eta \psi,$$

where k_j is the cost share of the jth input, η denotes the (cost) price elasticity of output, while ψ represents a function of the rate of return to scale such that $\psi = 1$ when this rate is constant.

In Hultkrantz's model, output will decrease with rising costs while the rate of return to scale is non-increasing. Even if an oil price hike would mean that capital and labor tended to replace energy the consequent downscaling of production could therefore lead to complementarity being registered with this kind of elasticity measure. The same evidently is true in regard to the substitution relation between electricity and oil.

The weak complementarity between oil and electricity in Lundgren's model of the iron and steel industry seems instead to be caused mainly by switches between different technologies. In the short-run version, with effective constraints on investment, the sign of the elasticity is reversed due to the fact that the electric arc furnace is then still a viable option. The own-price elasticity for oil recorded by Lundberg for the long-run version seems surprisingly large, which may at least partly be due to his probably unrealistic assumption of flexible furnace equipment, making possible a costless switch from oil to internally generated fuels like coke-oven and blast-furnace gas.

Structural change and energy use in the future

One of the reasons for measuring substitution possibilities is the need to gauge the future energy requirements in Swedish manufacturing. To discern future trends in industrial energy demand, one must study the dynamics of industrial investment and growth, analyzing the effects on specific energy use and tracing the changing branch composition.

That is the aim of the study by Ysander-Nordström making up Part III of this volume. The authors try to accomplish it by simulations on a dynamic macro model of the Swedish economy, incorporating a vintage approach to industrial capital, and a relatively detailed description of the different mechanisms for energy substitution. Many of these mechanisms have been modeled using the estimates of price elasticities derived by Dargay and Jansson.

Some of the most interesting results of this study are summarized in Table 3. For each form of energy the change of total use in manufacuring during the period 1980-2000 is recounted as the change in production volume multiplied first by the change in energy coefficients (structure being held constant) and then by the change in energy use structure (energy coefficients being kept constant).

We see that for total energy the "structural" effect is of the same magnitude as the change in specific energy usage. The same is true for total fuels and for electricity. The change in specific usage varies, however, between the fuels as to both sign and magnitude. While specific usage is halved in the case of oil it increases almost by half for coal and by some thirty percent for domestic fuels.

Some rather dramatic changes in the energy system are moreover expected to occur during the period. The closing down of nuclear reactors, beginning in the 90s, will mean an end to the "electricity glut" and will imply higher electricity prices, which can be expected to cause a certain slow down both of mechanization and electrification and of oil saving in manufacturing.

Table 3 Factors determining change of energy use in manufacturing, 1980-2000

	Relative change 2000/1980 in:			
	total production volume •	specific energy usage[a] •	use structure[b] =	energy use
Oil	1.65	0.52	0.93	0.79
Coal	1.65	1.46	0.86	2.07
Domestic fuel	1.65	1.29	0.83	1.77
Total fuel	1.65	0.90	0.87	1.29
Electricity	1.65	0.92	0.91	1.38
Total energy	1.65	0.90	0.88	1.31

[a] Weighted average of specific energy usage with 1980 production shares as weights.

[b] Weighted average of production shares with specific energy usage in 2000 as weights.

One way of summarizing the findings reported in Table 3 would be to note that half the total energy savings up till the turn of the century would be realized even if the average energy-efficiency remained unchanged within each manufacturing branch. Having worked our way through the maze of econometric estimates of substitution possibilities within the manufacturing branches, we thus come back to the conclusion already derived intuitively from postwar experience. Energy saving and energy economy are not just matters of public and private energy policy. They depend as much on economic development in general and on the rate of industrial restructuring in particular.

REFERENCES

Berndt, E.R. and B.C. Field (eds.), 1981, Modeling and Measuring Natural Resource Substitution. The Mit Press, Cambridge, Mass.

Berndt, E.R., C.J. Morrison and G.C. Watkins, 1981, "Dynamic Models of Energy Demand: An Assessment and Comparison"; in Berndt-Field (eds.). (Cf above).

Field, B.C. and P.G. Allen, 1981, "A General Measure for Output - Variable Input Demand Elasticities", American Journal of Agricultural Economics, 63.

Field, B.C. and E.R. Berndt, 1981, "An Introductory Review of Research on the Economics of Natural Resource Substitution", in Berndt-Field (eds.) (Cf above).

Kopp, R.J. and V.K. Smith, 1981, "Measuring the Prospects for Resource Substitution under Input and Technology Aggregations", in Berndt-Field (eds.). (Cf above).

McFadden, D., 1963, "Constant Elasticity of Substitution Production Functions", Review of Economic Studies, 30.

Parks, R.W., 1971, "Responsiveness of Factor Utilization in Swedish Manufacturing 1870-1950", Review of Economics and Statistics, 53.

Robinson, J., 1933, Economics of Imperfect Competition, London.

CHAPTER VI

Local Gorvernment and Economic Growth

One of the most striking features of the Swedish economy today is the growing dominance of the local government sector. Out of the national income almost one third is channelled through the budgets of local governments, who employ one fifth of the labor force. Local government spending has been outrunning the GNP with a growing margin, doubling its share over the last twenty years.

Organizational power has grown with the money. In the postwar period there has been a gradual concentration of the decision-making process. While the number of local government has decreased to about one-fourth in the last fifteen years—24 counties and 277 municipalities right now—some of the most expansive areas of service production, like mental health care and secondary schools, have been taken over from the state by local government.

The sphere of responsibility of local governments has, therefore, widened significantly both in form and in content. Individual counties and municipalities today have much greater possibilities of independent long-term planning and procurement and of negotiating with the central government and with large corporations. Along with the successive diversification of the services supplied there has also been a widening of their area of responsibility through their increased participation in distribution and stabilization policies.

Throughout this expansion the local governments have retained a degree of financial independence of central government, which is rather high by West European standards. Of their total gross expenditures only about one-fourth is paid by state grants, while local taxes make up for 45 %, fees and user charges 20 %, with loans and capital income making up the remaining 10 %. For high-lighting the degree of tax-financing of the local services, another way of calculating may however be more relevant. If one includes only net profits—or losses—of public utilities, net new borrowing and net transfers from the state (i.e. subtracting taxes and fees paid by local governments to central government) another financial picture emerges. Of the total local government spending directed toward the private sector about 90 % was paid by taxes of which only one sixth was channeled through the state budget. The remaining 10 % was made up of some 7 % for fees and only about 3 % of new loans.

The growing importance of local governments as an independent force in the Swedish economy has motivated an econometric study within IUI, from which the historical data and perspectives presented in the following are extracted.

A Long-Term Perspective

Figure 7 represents total local government expenditure as a percentage of GNP at factor prices. The figure shows how a relatively slow rate of growth during the 1920's, 30's and 40's was replaced, from the beginning of the 1950's, by a dramatic expansion that is still continuing.

Figure 8 shows how local government expenditures were distributed among spheres of activity in the years 1913, 1953 and 1977. The relative development pattern changed from the period between the two world wars to the postwar period. While, for example, in the earlier period education and industrial activities tended to increase their share of total local government expenditure, they now account for a successively smaller part of the budget. Provision for housing, however, shows a development in the opposite direction, while medical and health care has, during the entire period, increased its share of total expenditure.

Figure 9 represents the three largest and fastest growing categories of local government expenditure—education, medical and health services and social services. The latter two categories are the only ones that, in terms of averages for the whole period, have expanded faster than total expenditure.

Medical and health services have been growing fast since the 1930's, while the growth rate of social service expenditures has substantially exceeded that of total expenditure only during the beginning of the 1930's (the years of the Great Depression) and the 1970's. Expenditures on education have throughout followed the same pattern of growth as total expenditure.

Figure 10 presents in a corresponding manner the development of local government expenditures on those spheres of activity that are, in a wider sense, inter-connected with physical planning. During the inter-war period, expenditure on housing increased at a rate slower than that of total expenditure but has made up for this by growing faster after World War II.

Expenditures on roads and highways showed similar development until the beginning of the 1970's but have since been given a low priority both by central and local government. Finally, expenditures on justice and law enforcement, municipal planning and general administration show a relatively irregular development where, for example, stagnation during the latter part of the 60's gave way to an accelerated expansion during the 70's, particularly marked in the case of expenditure on general administration. Part of the accelerated increase of this group of expenditures is probably explained by the numerous mergers between municipalities in the early 70's.

Figure 7. *Local government expenditure as percentage of GNP, 1913—1980*

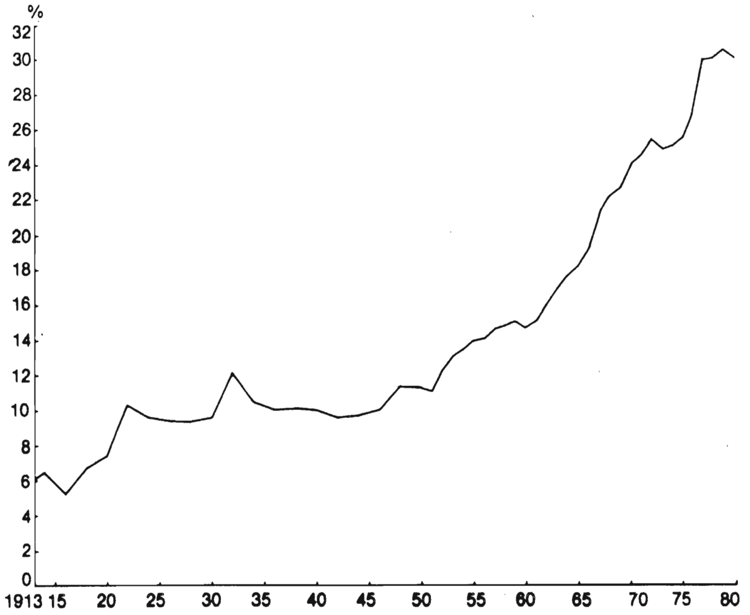
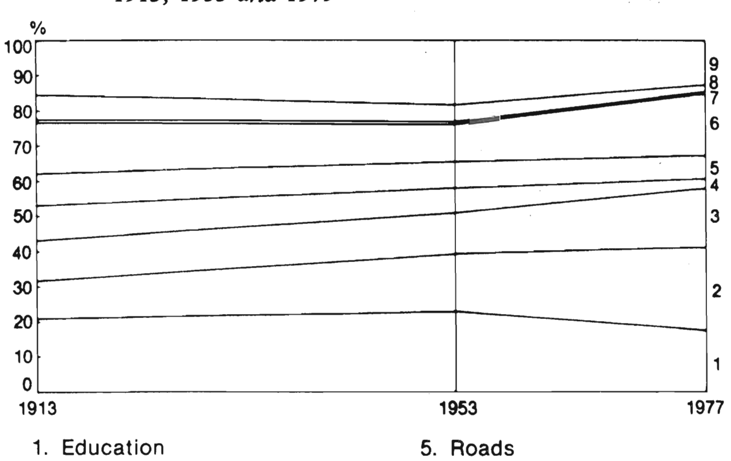


Figure 8. *Distribution of local government expenditures on types of activity, 1913, 1953 and 1979*



1. Education
2. Medical care
3. Social welfare
4. Justice, administration and municipal planning
5. Roads
6. Housing
7. Fire protection
8. Churches
9. Industrial activities

Figure 9. *Local government expenditure for education, medical care and social services, 1913—1977*
Index 1913 = 100

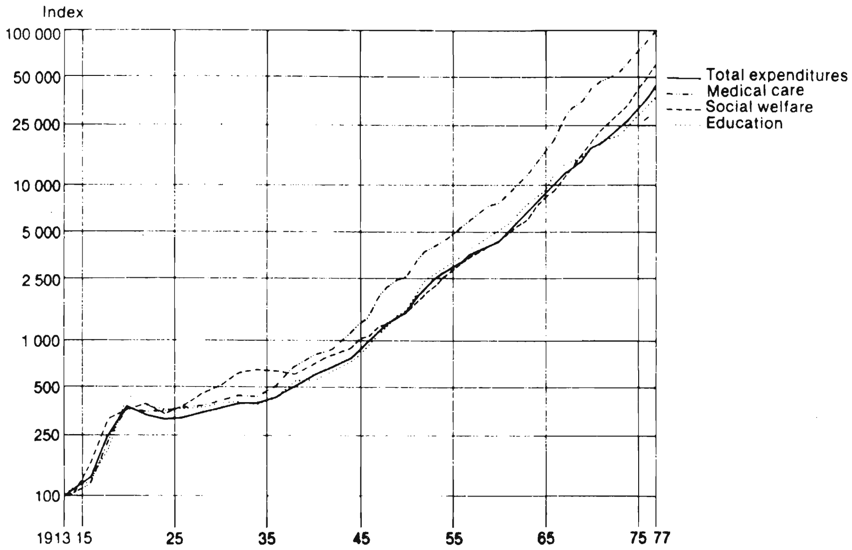
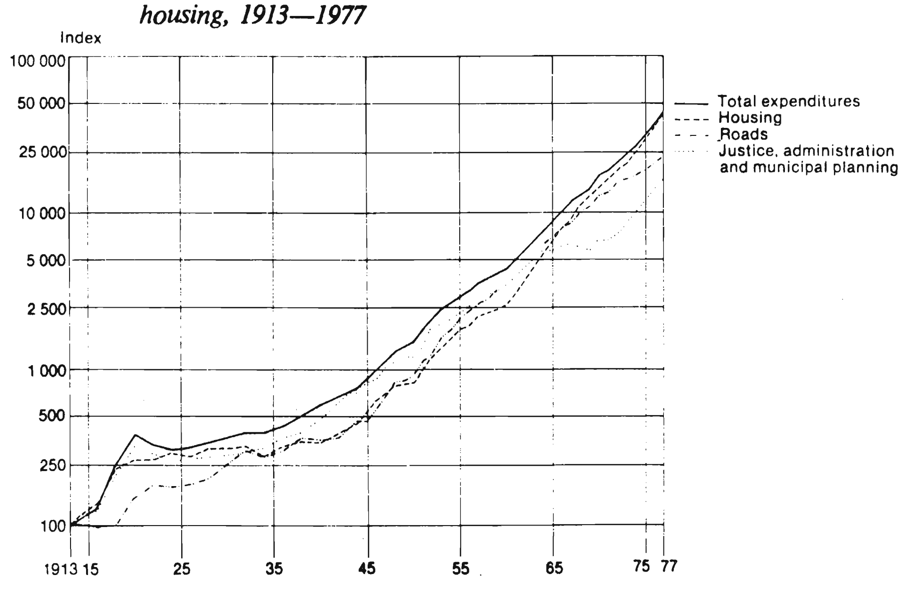


Figure 10. *Local government expenditure for justice and law enforcement, general administration and municipal planning, road services and housing, 1913—1977*



The Swedish Welfare Strategy

The restructuring of the Swedish economy in postwar years has been rapid although not exceptional compared to other western countries. Over the thirty years since 1950 agricultural employment has been drastically reduced and corresponds today to less than 5 % of the total labor force. The matching increase has occurred in the services, particularly in the public services, which doubled their share of GNP and trebled their employment share. The major part of this expansion took place within the local government sphere—in education, medical care and social welfare. Manufacturing industry meanwhile kept its share both of GNP and employment relatively unchanged. The enlarged public service provision was almost entirely paid for by taxes, which trebled relative to GNP. It was again local governments that were responsible for the major part of the tax increases.

What has given the Swedish welfare state a characteristic profile of its own is not only the size of the public budgets, although it tops the list of international statistics in this respect, with roughly two-thirds of all income being channeled through public budgets. There also is what could be called a characteristic Swedish welfare strategy concerned with the ways of using public budgets for redistribution purposes.

On the spending side there are two major alternatives as to how to make social services freely available. Government can do it by subsidized insurance schemes or by direct income transfers to cover the necessary costs, leaving at least part of actual choice of service procurement and use with the individual. Alternatively, the government may assume monopolistic responsibility for service supply and distribute the service free of charge but constrain and regulate the access. Sweden has in postwar years very decisively chosen the latter course. Compared with major West European countries like Italy and France both the postwar growth and the level reached of public transfers relative to total income have been rather moderate in Sweden. It is instead the public price subsidies, in particular the tax-financing of the social services, that have grown exceptionally fast and now dominate the public budgets. This choice of strategy has at least partly been induced by the prevalent notion of free medical care and social welfare services having a more substantial redistributive impact than any alternative payment schemes.

The importance in Swedish public budgets of price subsidies in this general sense is demonstrated by Figure 11, where total public spending has been divided according to the "mode of distribution" into expenditures for traditional collective goods like defense, justice and central administration, price subsidies dominated by local government provisions and finally income transfers of which a major part are social insurance payments from central government.

A marked concern for redistribution also characterizes the Swedish "taxing strategy". In all countries the importance of indirect taxation has grown in the postwar period along with income transfers—partly due to the enlarged coverage of social insurance. As shown in Figure 12 the personal income tax in Sweden, however, has increased in step with total taxes—in contrast to the changing tax structure of i.a. France, Italy and West Germany. The central government part of the income is highly progressive, although it is hard to know how much of actual redistribution that is effected. We can also see from the figure that the households nowadays receive almost as much transfer money from government as they pay out in income taxes.

Figure 11. *Mode of distribution of public spending in Sweden 1950—1980*

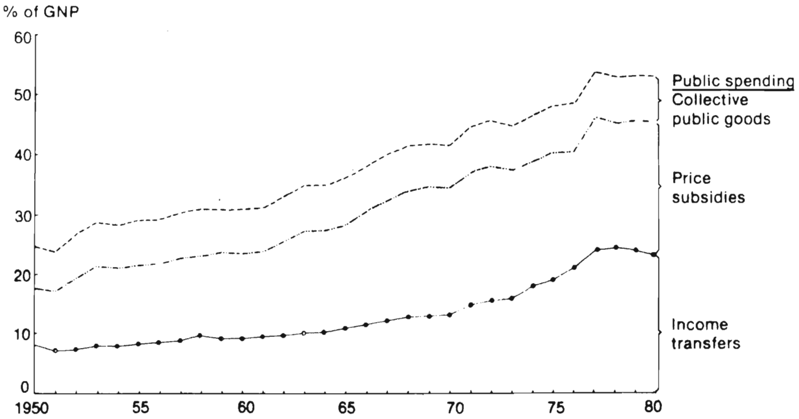
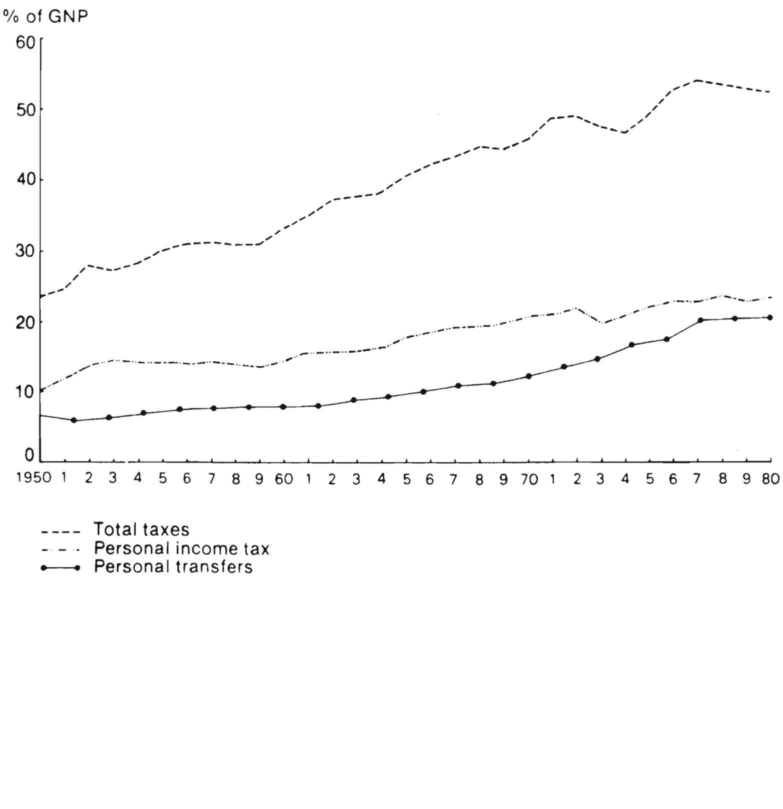


Figure 12. *The taxing strategy*



Is Local Government Spending Out of Control?

In recent years there has been a rising concern in Sweden about the development of local government expenditures. The rapid expansion of local government services has often been blamed not only for pushing up the tax scales and with that inducing various disruptive tendencies towards tax evasion and tax adjustment. It has also been suggested that local government competition in the labor market has contributed significantly to wage inflation and recruitment problems within Swedish industry. The accelerating cost increases in the social services during the latter part of the 70's have been interpreted by many as signs of a falling productivity due to agencies creating too many new service jobs as part of the efforts to keep down open unemployment.

Looking ahead into the 80's one of the few things we can be fairly certain about is that local government expansion cannot be allowed to continue at the rate established during the 60's and 70's. We have neither the goods nor the people to sustain that kind of growth. Even in absolute amounts the annual increases in local government resources will have to be somewhat reduced if we want to get rid of our external payment problems before the 90's and avoid having in the meantime to lower real net wages and private standards. However, making the municipalities change fast enough into a slower growth-track in spite of good liquidity and rather rigid long-term plans may well prove to be one of the crucial economic problems in the next few years.

Researching the Problems

It is against this background that IUI has started an econometric study of local government behavior in Sweden. An explanatory model of local government spending—split up into 16 different categories—and taxing decisions has been estimated on data from the 60's and 70's. Based on this model, various specific problems are now being researched. Service production data are tested to detect possible changes in production "techniques". The effectiveness of central government grant policy in various areas of service production will be measured.[1] By integrating the local government model into a large growth model for the total economy we are able to study interactions between local government expansion and growth in other sectors.[2] Finally the model has been used for projection of actual local government spending and taxing into the 80's.[3]

[1] A first attempt of a closer look at the effects of grants on local government expenditures and employment has been presented in: E.M. Gramlich and B.-C. Ysander, *Relief Work and Grant Displacement in Sweden*, IUI Working Paper No. 30, 1980.

[2] One such projection was documented and discussed in: B.-C. Ysander, "Offentlig ekonomi i tillväxt" (Public Service Growth), Chapter 8 in *Att välja 80-tal* (IUI Medium Term Survey 1979), IUI, 1979.

[3] Cf. i.a. T. Nordström and B.-C. Ysander, "Offentlig service och industriell tillväxt" (Public Service Supply and Industrial Growth), IUI Research Report No. 11, 1980.

CHAPTER VII

Public Policy Evaluation in Sweden[1]

Auditing the Swedish Welfare Economy

Auditing a market economy normally means making sure that effective competition is maintained. Auditing a welfare state is a much more complex task. It means i.a. evaluating the efficiency of government monopolies by the use of hypothetical market analogies. At the same time, the equity considerations of the welfare state have to be observed. The outcome of auditing will depend on the choice of criteria for comparison – alternative government policies or market solutions, alternative taxes, lumpsum transfers etc. The more dominant and all-embracing the public sector becomes and the more ambitious the redistribution policies are, the more confused the Government objective function will be and the more difficult it becomes to do the auditing work. But the reason for doing it will be all the more pressing.

There are several pressing reasons for Sweden to be particularly concerned about the evaluation of public policy.

Tax rates and public spending shares are the highest in the world. More than 70 percent of total income is channeled through public budgets rather than through markets.

Compared to other West European countries the Swedish welfare strategy is based on the provision of free public services, implying both relatively more public employment and long-term and inflexible commitments of public funds.

The rapidly increasing Swedish public budgets are dominated by expenditure used for price subsidies in general, and public consumption in particular. The share of income for collective security has remained more or less constant around 10 percent over the whole postwar period. The dramatic

[1] Revised and shortened version of a lecture given at Colloque International "L'Evaluation des Politiques Publiques", Paris, 15-16 décembre 1983, published as "L'évaluation des politiques publiques en Suède", in Nioche, J.-P. and Poinsard, R. (eds.), *L'évaluation des politiques publiques,* Economica, Paris, 1984, and in IUI Booklet No. 215, Stockholm.

expansion of the public budget share is entirely due to social security expenditure, which has almost tripled its share during the last 30 years.

Most social security expenditures can be said to be ultimately concerned with redistributing real income. This may take the form of insuring against social and economic risks, redistributing resources over the individual's lifetime or shifting the levels of life income prospects between individuals. This means that policies have been focused on the distributive effects. One reason why policy makers so far have often been unappreciative towards attempts at economic policy evaluations may indeed be their preoccupation with feasible redistributions. Economists, on the other hand, often treat redistribution as a side-issue, or a restriction on their main concern with efficiency and/or macroeconomic stabilization.

Policy ambitions are mirrored by the perceptions or models of economic reality used in public economic analysis. One problem is that these perceptions have changed as a consequence of the economic events of the 70s.

Policy Evaluations in a Swedish Context[1]

Some kind of policy evaluation normally precedes policy making. The policy cycle begins with *policy analysis* – the ex ante evaluation of options on which the *policy decision* is based. Then comes implementation and finally ex post *evaluation* – the theme of this paper – hopefully operating as a learning experience for the next round of policy making (Edlund, and others, 1981, and Wildavsky, 1979).

What one ultimately wants to evaluate is the *social effectiveness* of the implemented policy, i.e. its effects on the welfare of individuals and groups in the community. In most cases this is the same as its impact on the size and distribution of real income.

A useful distinction can be made between, on the one hand, *policy effectiveness* and, on the other hand, *management efficiency*. The second measures the efficiency in implementing policies. A good deal can be learned by simply looking inside Government offices. Evaluating policy effectiveness, on the other hand, almost invariably requires "field studies" of the policy impact on private individuals and organizations (Farell, 1957, and Førsund, Lovell, and Schmidt, 1980).[2]

[1] For an alternative resumé of the Swedish experience in public policy evaluation cf. Premfors, 1984.

[2] The reader will notice that, contrary to the practice among business economists, we here use effectiveness as a broader concept than efficiency, encompassing also distributional considerations. The simplified distinction used above thus disregards the fact that implementation decision on the management level may also have important distributional consequences. For an extensive discussion of efficiency concepts and their applications to public administration, cf. Jackson, 1982.

The reader will notice that, contrary to the practice among business economists, we here use effectiveness as a broader concept than efficiency, encompassing also distributional considerations. The simplified distinction used above thus disregards the fact that implementation decision on the management level may also have important distributional consequences. For an extensive discussion of efficiency concepts and their applications to public administration, cf. Jackson, 1982.

In evaluating efficiency in the private economy, economists often argue that welfare losses due to misallocation are negligible compared to the losses due to inefficient resource use within each line of production (Leibenstein, 1966). There are reasons to assume that the opposite applies to the public economy. Apart from distortions due to taxes and subsidies, there are the problems involved in "filtering" preferences through a representative democracy and its bureaucratic machinery. Intuitively, one would therefore expect the "non-market failures" to be far greater than the "market failures".

At least in Sweden, public opinion tends to regard the problem of inefficient public administration as limited compared to the risks of ineffective policy choices. Inefficient public bureaucracy therefore appears as less of a problem in Sweden than in most other countries. Many factors have contributed to this, notably a long tradition of disciplined and incorrupt bureaucracy. An overgrowth of central administration has not yet occurred. Moreover, the public sector is mainly associated with health and education – status goods in expanding demand. Compared to most other countries public policy in Sweden is also more decentralized. The relatively independent local authorities, municipalities and counties are responsible for more than two thirds of all public consumption (Ysander-Murray, 1983). Even central Government power is decentralized; policies are mainly executed, and often also initiated, by independent national agencies.

Although the number of domestic policy-issues in Sweden is comparable to that of a larger country, a small country has less resources for specialized policy evaluation. Decentralization has often provided an excuse for not even trying. Conventional wisdom among politicians is that decentralized decision-making is a substitute for policy evaluation. Public attention to policies is more immediate, when decisions are made "closer to the market".

There are two additional features of Swedish postwar politics, that have tended to lessen the interest in evaluating policies.

Many of the political institutions – like Government Committees – have been designed to produce consensus decisions. At any time there are 200-300 of these committees at work, with an average lifetime of 3-4 years. [1] They

[1] 206 Govt. Commissions were at work in the autumn 1983. Efforts are being made to speed up the investigative process, aiming at a maximum lifetime of 2 years.

are composed of MPs and representatives of different interest groups. Their task is to prepare – and negotiate – major policy changes and new legislation. Committees contract outside experts to review the past or develop new options, but are usually narrowly constrained by Government directives. Consensus politics mean that decision-making takes time. It might even include evaluation of past policies. But once consensus decisions are taken, interest in reappraisal tends to vanish.

Major policy decisions usually represent a heavy investment in terms of political credit. A certain amount of indoctrination is usually required to ensure support and acceptance. Hence, enthusiasm among responsible parties for later checking policy arguments against facts is normally lacking.[1]

The very rapid economic growth in the 50s, 60s and early 70s furthermore focused political interest on policy expansion and incremental change rather than on policy restructuring and alternative options. This necessarily limited the possibilities of evaluation by narrowing the range of "experimental variation" in the available data.

From Program Evaluation to Problem Reappraisal

Before 1960 the Government made no systematic policy evaluation. Evaluations – if they occurred – were initiated by some Government Committees to develop arguments for new legislation. The monitoring of public administration by the National Audit Bureau was limited to safe-guarding the interest of fiscal regularity and public accountability – what is nowadays often termed *compliance auditing*.

The period from 1960 to the late 70s witnessed an expansion in *program evaluation*. Both the economy and the public sector were growing rapidly, particularly local authorities, which expanded almost twice as fast as GNP. How to organize a trebling of university students and yet accommodate an even faster growth of adult studies was a typical concern of Government. Health and welfare services and pension schemes had to be prepared for a doubling of the number of old-age people. Public child-care capacity "had to" double to facilitate female labor participation needed to replenish an overheated labor market. Ambitions expanded and began to include better labor market matching and retraining, an "improved" regional balance of manufacturing investments, etc. Not least important was the need to reorganize public administration to cope with all new tasks.

Contrary to what happened in the Anglo-Saxon countries, program budgeting, PPBS[1], was developed and introduced in Sweden not as an in-

[1] Tarschys provides a stimulating discussion of the waxing and waning of political interest in policy evaluation during different phases of policy-making (see Tarschys, 1983).

[2] Planning – Programming – Budgeting – Systems.

strument for central Government policy making but as a way of decentralizing the administration to increase efficiency at the agency level. Program budgeting became an accepted routine for an increasing number of agencies. A unified scheme for program cost accounting and internal program reviews were gradually introduced in the agencies. The National Audit Bureau transferred its resources to selective checks on agency decision-making and reviews of programs. However, the risk of "in-built" expansion, always inherent in the PPBS approach, still remains a problem, although endogenous expansion has been curbed through enforced plans for program reductions at agency level (National Audit Bureau, 1983a and b, SOU 1979:61).

At cabinet level, ad hoc Government Committees were subjected to competition when several large ministries initiated their own R&D committees, with a semi-independent and semi-permanent status, staffed with both experts and civil servants. These committees were authorized to monitor and initiate policy research. Most policy evaluation at the time in fact took place in these committees. The record of serious policy evaluation, however, is far from impressive. Methods were crude and efforts low keyed. A few pioneering attempts at statistical analysis of program effects can be noted from the early 70s (SOU 1974:29, Björklund, A., 1981, and Kjellman, S., 1975). These efforts, however, cannot match the steady outflow of studies and the rapid development of statistical evaluation methods achieved in the United States ever since the negative income tax experiments (Guttentag, Struening, 1975, and Premfors, 1984). Some experiments were carried out locally in the social welfare fields, but these experiments were seldom used as a basis for a full scale evaluation. Economic evaluations of medicines and medical treatments did, however, become increasingly frequent. A common problem which we have as yet done little to solve concerns the poor availability of relevant and reliable panel data. The lack of good data is also the excuse often used to explain the very small amount of evaluation work in the field of taxes and transfers.

Within university education, however, policy changes and reform evaluations have been frequent in the postwar years (Edlund and others, 1981; Neave and Jenkinson, 1983; and Premfors, 1983).

After 1976, the long period of postwar prosperity was succeeded by industrial stagnation. A bourgeois Government was elected after more than forty years of socialist hegemony. Policy evaluation from now on could best be characterized as problem reappraisals. The mounting economic and financial problems, and the frequent changes in Government made it both possible and necessary to reconsider basic policies and conventional wisdom (SOU 1979:61). The political consensus was breaking apart and the climate of opinion was undergoing drastic changes.

This led not only to a heightened interest in policy evaluation but also to a change of direction of the evaluation work. From having been mainly "program-oriented", evaluation work has become increasingly "problem-oriented". Instead of starting at the top level with an individual program like labor market retraining, and following it down the line to its final execution, trying to measure its differential impact on individuals and firms, the tendency now is to go the other way around. One begins by studying the total impact on a specific target group, such as children in day care.

At the management level, this has meant new tasks for the National Audit Bureau, which is now allowed not only to look at individual agencies and programs, but also to reappraise the efficiency of program and agency structure. The National Audit Bureau has established routines for computing total public transfers for various types of households and firms, and is studying the effects of diverse licensing laws and of deregulation measures in progress.

At the policy-making level there is a new interest in evaluating whole policy systems by comparing them with radically different alternatives. The Treasury has recently established its own R&D-committee, using it as a sounding-board for new policy options. Its interest extends to the appraising of new transfer structures and new models for social insurance.

Table 1 presents the various modes of public policy evaluation mentioned above in a summary fashion. In terms of this table Swedish development

Table 1 *Different modes of public policy evaluation*[a]

Object of study	Method of approach	Program-oriented study of incremental change	Problem-oriented study of intra-marginal change
Management regularity		Compliance auditing	Studies of budgetary control systems
Management efficiency		Management auditing. Cost-effectiveness studies	Studies of bureaucratic systems
Policy effectiveness		Effectiveness auditing	Social welfare studies. Total impact studies of public policy.
(Allocative efficiency, distributional effects)		Program evaluation Cost-benefit analysis	Studies of alternative modes of financing and distributing public services and insurance.

[a] For a more elaborate classification scheme for public evaluation cf. e.g. Ahonen, 1983.

since 1960 can be characterized as a shift of emphasis "downwards" – from management regularity to policy effectiveness – and "to the left" – from program-orientation to problem-orientation.

However beneficial reappraisals are, they cannot replace the painstaking work of analyzing program impacts. Unfortunately, such work has not progressed in late years, and no effort has been made to build the necessary foundations in terms of good panel data and trained analysts. In the case of labor market policy and social welfare policy very little has been done to continue the statistical analysis initiated in the early 70s.

Only in one field has there been a fast expansion of evaluation work during recent years, viz. energy policy. A deluge of energy research funding has been channeled into policy and project evaluation work (Andersson-Bohm, 1981; Vedung, 1982; and Ysander, 1983).

Development in policy analysis during the last decade has also been disappointing. Benefit-cost analysis has, so far, become an administrative routine only at the National Road Agency. Good benefit-cost work on public projects is rare. The same is true for policy analysis using large scale simulation models, although an increasing interest has been noticed in the last few years (Carlsson-Bergholm-Lindberg, 1981; Vedung, 1982; and Ysander, 1983). There is of course a close connection between developments in policy analysis and policy evaluation. Evaluating policy means evaluating a social experiment. It is then important to know, by policy analysis, the expected consequences of the experiment.

The Evaluation Bureaucracy

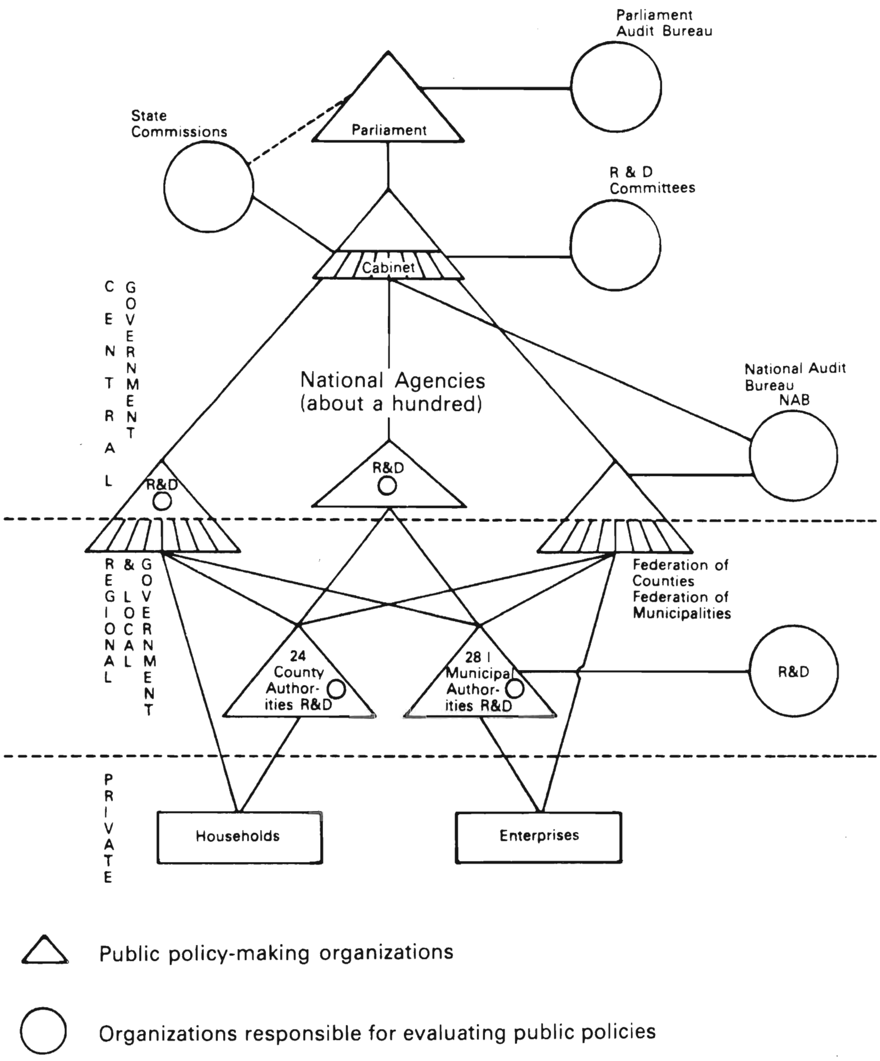
The organization of Swedish evaluation work is exhibited in Figure 1.

Parliament and Cabinet

The Parliament Audit Bureau is quite small and subordinated to a board of MPs. Its size and the political monitoring of its analysis have, so far, tended to reduce its role. This also reflects the weakened position of Parliament vis-à-vis Government during the last half-century. Frequent changes of government and the precarious parliamentary balance in recent years have not really changed that situation.

The traditional and dominant vehicles for policy evaluations in Sweden are the Government committees. Government directives and the tight time schedules of committee work narrowly limit the scientific ambitions of the evaluations. The most important function of the committee is to prepare the way for a consensus decision in Parliament. Hence, the expert arguments are often used more as political ammunition than as an objective support of decisions (Premfors, 1983).

Figure 1 *The Swedish organizational structure*



The second important source of policy evaluations at cabinet level are the temporary R&D committees, set up by ministries like Justice (BRÅ), Labor (EFA), Industry (ERU), Social Welfare (DSF) and Treasury (ESO).

National Agencies

A Swedish minister of state has a power position very different from, say, his French colleague. All cabinet decisions are taken collectively, and the minister's own staff rarely exceeds 40-100 people. Most executive work is handled by the associated national agencies in accordance with a 300-year-old tradition. The national agencies enjoy a high degree of autonomy and protection from direct ministerial intervention. One hundred such national agencies and a couple of hundred minor national organs are responsible for current resource allocation and for the issuing of regulations and directives to local authorities and private organizations. Some of the major agencies also have a large regional organization.

Regional and Local Agencies

Even though relatively autonomous by international standards, the Swedish counties and municipalities are regulated by the state. Around 70 percent of local expenditures are somehow regulated and some 30 percent of these regulated expenditures are, on average, paid by the state. Comprehensive schools and highschools in the municipalities are subjected to particularly heavy subsidies and regulations. The same goes for medical services, which are the main responsibility of the counties.

Internal auditing and reviewing within local authorities therefore emphasizes management effectiveness. Since "municipal mergers" in the early 70s reduced the number of units to a third, better and more unified systems for cost accounting and financial management have been organized.

A number of counties, and some of the major municipalities in metropolitan areas do, however, have their own R&D units, for planning large investment projects and for monitoring labor market flows.

Decentralization and Fragmentation

The slow progress of policy evaluation in Sweden is best explained by its decentralized and fragmented organization.

The *decentralized* structure of Swedish Government has eased the political pressure for central government monitoring. Evaluations of policies, for which responsibility rests with the local authorities, may e.g. often be considered not only less urgent but even politically unsuitable for organs of central Government. Attempts in postwar years to have interest groups or client representatives directly involved at different levels of the National Agencies have been seen as a vehicle for faster and more direct feedback.

There is always a political tug-of-war between, on the one hand, the groups clamoring for centralized regulation and resources to protect their interests or the equality of standards and, on the other hand, the more general pressure for decentralization and deregulation in the name of efficiency and freedom (Tarschys, 1975 and 1983b). The last decade of Swedish politics has witnessed strong swings in both directions with restrictive labor market legislation and heavy industrial subsidies (Carlsson-Bergholm-Lindberg,

1981) on the one hand, and on the other, a flow of actual or proposed deregulation measures.

The Swedish organization of policy evaluation is also very *fragmented.* A major part of evaluation work is initiated and financed by temporary government committees and commissions with very limited budgets, tight time schedules and narrow political directives. In most cases they have been set up to investigate a specific proposal. Evaluations of past policies therefore tend to be not only limited but also superficial, relying in most cases on a review of already documented experiences. No individual commission has the right, the resources or the patience to conduct a full-scale statistical post mortem on important policy choices in the past. Neither will they plan their proposals in order to facilitate later evaluation (Premfors, 1983b). Even though references to evaluation requirements have become frequent in government policy documents, so far these requirements have been more related to management efficiency than to policy efficiency.[1] The fragmented organization also makes evaluation difficult in another way. There are increasing returns to scale in evaluation work in the sense that everybody can benefit from the production of good economic and social data, and from the building up of a common body of expertise within Government. Individual committees etc. can reap the benefits of such common resources but usually they can do little alone to produce them.

References

Ahonen, P., 1983, *Public Policy Evaluation as Discourse,* Helsinki.

Andersson, R. – Bohm, P., 1981, *Samhällsekonomisk utvärdering av energiprojekt* (Economic Evaluation of Energy Projects).

Björklund, A., 1981, *Studies in the Dynamics of Unemployment,* EFI, Stockholm School of Economics.

Bohm, P., 1972, "Estimating Demand for Public Goods: An Experiment", *European Economic Review* No. 2.

Carlsson, B. – Bergholm, F. – Lindberg, T., 1981, *Industristödspolitiken och dess inverkan på samhällsekonomin* (The macroeconomic effect of industrial subsidies), IUI, Stockholm. For a summary in English, see Carlsson, "Industrial Subsidies in Sweden: Simulations on a Micro-to-Macro Model" in *Microeconometrics, IUI Yearbook, 1982-83,* Stockholm 1983.

Edlund, C., et al., 1981, *Utvärderingsforskning* (Evaluation Research), Studentlitteratur, Lund.

Eliasson, G, – Ysander, B.-C., 1983, "Problems of Maintaining Efficiency under Political Pressure", in Hindley, B. (ed.), *State Investment Companies in Western Europe,* McMillan, London.

[1] This conclusion has been drawn from a survey of the ministries, carried out in September 1983 by the Committee for Studies on Public Finance (ESO) on behalf of the author.

Farell, M.J., 1957, "The Measurement of Productive Efficiency". *Journal of the Royal Statistical Society,* Vol. 120, Series A.

Førsund, F.R. - Lovell, C.A.K. - Schmidt, P., 1980, "A Survey of Frontier Production Functions and of their Relationship to Efficiency Measurement", *Journal of Econometrics,* No. 13.

Gramlich, E.M. - Ysander, B.-C., 1981, "Relief Work and Grant Displacement in Sweden", in Eliasson-Holmlund-Stafford (eds.), *Studies in Labor Market Behavior: Sweden and the United States,* IUI, Stockholm.

Guttentag, M. - Stuening, E.I. (eds.), 1975, *Handbook of Evaluation Research I-II,* Sage, Beverly Hills.

Jackson, P.M., 1982, *Political Economy of Bureaucracy,* Philip Allan, Oxford.

Kjellman, S., 1975, *Arbetsmarknadsutbildningens effekter på inkomstfördelning och resursanvändning* (The Distribution and Resource Allocation Impact of Labor Market Training), Report No. 9, Parliamentary Audit Office.

Leibenstein, H., 1966, "Allocative Efficiency vs. X-Efficiency", *The American Economic Review,* Vol, 56, No. 3.

Markstedt, L., 1982, *Abolishing Rules and Regulations,* National Audit Bureau, Stockholm, 1983-11-17.

May, J. - Wildavsky, A. (eds.), 1978, *The Policy Cycle,* London.

Murray, R., 1981, *Kommunernas roll i den offentliga sektorn* (The Role of Local Governments in the Public Sector), Mimeo, IUI, Stockholm.

National Audit Bureau, 1983a, *Application of Integrated Budgeting System in Sweden,* Report 1983-08-25, Stockholm.

National Audit Bureau, 1983b, *New Development Trends within Effectiveness Auditing in Sweden,* Mimeo, Stockholm.

Neave, G. - Jenkinson, S., 1983, *Research on Higher Education in Sweden,* Almqvist & Wiksell, Stockholm.

Nordström, T. - Ysander, B.-C., 1980, *Offentlig service och industriell tillväxt* (Public Consumption and Industrial Growth), IUI Research Report No. 11, Stockholm.

Premfors, R., 1981, *Genomförande och utvärdering av offentlig politik* (Implementation and Evaluation of Public Policy in Sweden), Report No. 12, Department of Political Science, University of Stockholm.

Premfors, R., 1983a, *Implementation Strategies in Higher Education,* Report No. 26, Department of Political Science, University of Stockholm.

Premfors R., 1983b, "Social Research and Govermental Commissions in Sweden", *American Behavioral Scientist,* Vol. 26, No. 5.

Premfors, R., 1984, *Policyanalys i Sverige* (Policy Analysis in Sweden), Report No. 30, Department of Political Science, University of Stockholm.

SOU 1974:29, *Att utvärdera arbetsmarknadspolitik* (The Evaluation of Labor Market Policy), 1974, Government Printing Office, Stockholm.

SOU 1977:20, *Kommunernas ekonomi 1975-85* (The Economy of Local Governments), Government Printing Office, Stockholm.

SOU 1979:61, *Policy Innovation through Policy Reappraisal,* Government Printing Office, Stockholm.

Stromsdorfer, E.W. - Farkas. G. (eds.), 1980, *Evaluation Studies Review Annual,* Vol. 5, Sage Publications, Beverly Hills.

Tarschys, D, 1975, "The Growth of Public Expenditure: Nine Modes of Explanation", *Scandinavian Political Studies,* Vol. 10.

Tarschys, D, 1983a, "Att bekämpa krångel och onödig byråkrati" (To Fight Unnecessary Bureaucracy), *Nordisk Administrativ Tidskrift,* Vol. 64, No. 1.

Tarschys, D., 1983b, "Fluctuations in the Political Demand for Policy Analysis", in *Evaluating the Welfare State: Social and Political Perspectives,* Academic Press, Stockholm.

Vedung, E, 1982, *Energipolitiska utvärderingar 1973-82* (Evaluation of Energy Policy 1973-82), Report No. 52, Energy Research Commission, Stockholm.

Wildavsky, A., 1979, *Speaking Truth to Power: The Art and Craft of Policy Analysis,* Little Brown & Co., Boston.

Ysander, B.-C., 1979, "Offentlig ekonomi i tillväxt" (The Public Sector in a Growing Economy), in *Att välja 80-tal* (Choosing the 80s), IUI, Stockholm.

Ysander, B.-C., 1982, "Public Budgets in Sweden", *Public Budgeting and Finance,* Vol. 2, No. 3.

Ysander, B.-C. – Robinson, A., 1982, "The Inflexibility of Contemporary Budgets" and "Re-establishing Budgetary Flexibility", *Public Budgeting and Finance,* Vol. 2, No. 3.

Ysander, B.-C. (ed.), 1983, *Energy and Economic Adjustment,* IUI, Stockholm.

Ysander, B.-C. – Murray, R., 1983, *Kontrollen av kommunerna* (Central Control of Local Authorities), IUI Research Report No. 18, Stockholm.

Ysander, B.-C. – Nordström, T., 1984, "Local Authorities, Economic Stability and the Efficiency of Fiscal Policy" in Gramlich, E.M, - Ysander, B.-C. (eds.),*Control of Local Government,* IUI, Stockholm.

Selected Bibliography

ECONOMIC THEORY AND ECONOMETRICS

A. Methods of Economic Analysis

Articles

[A1] "Truth and meaning of economic postulates", in *Money, Growth and Methodology* (ed. Hegeland, H.), Lund 1961.

[A2] Review of J. Kornai: 'Antiequilibrium' (together with C.-H. Siven), *The Swedish Journal of Economics*, Vol. 75, No. 3, 1973.

[A3] "Guy Arvidsson. A short appraisal of his contribution to economics", *The Swedish Journal of Economics*, Vol. 76, No. 3, 1974.

[A4] Review of K. Arrow: 'The Limits of Organization', *The Swedish Journal of Economics*, Vol. 77, No. 2, 1975.

[A5] "Comment to Richard Day's paper: A Note on Endogenous Preferences and Adaptive Economizing" and "Comment to Carl Ch. von Weizsäcker's paper: Rights and Relations in Modern Economic Theory" – both in *Dynamics of Market Economies* (eds. Day, R. – Eliasson, G.), IUI and North Holland, 1986.

Working Papers

[A6] *Reperkussionsanalysen. En studie kring Frischs reperkussionsmodell.* (Repercussion analysis – a study on Frisch's model of repercussion). Memorandum, Dep. of Economics, Lund University, 1952.

[A7] *Tre teorikriterier och deras tillämpning på nationalekonomisk metodik* (Three theoretical criteria and their application in economics). Memorandum No. 23, Dep. of Economics, Gothenburg University, 1972.

B Resource Allocation, General Equilibrium and Welfare Theory

Books

[B1] *Resursfördelning.* Lärobok i resursfördelning med särskild tonvikt vid välfärdsteori och allokeringsproblem inom den offentliga sektorn (Resource Allocation; textbook on the economics of resource allocation with special emphasis on the theory of welfare and application as applied to the public sector), Utbildningsförlaget, 1971.

[B2] *Förvaltningsekonomiska problem* (Problems in public management) (ed.), Jurist- och Samhällsvetarförbundets förlag, 1972.

Articles

[B3] "Förvaltningsekonomiska problem" (Problems in public management), in the anthology [B2] above.

[B4] "Ransoneringskonstens ABC" (The ABC of economic rationing), *Ekonomisk Debatt*, Vol. 2, No. 1, 1974.

[B5] "The Meaning of Human Capital", in *Learning and Earning* (Bowman, M.J., Sohlman, Å. and Ysander, B.-C.), Liber, Stockholm 1978.

[B6] "Från blandekonomi till förhandlingsekonomi" (From the mixed economy to the negotiation economy), in *Svensk ekonomisk politik* (eds. Lundberg, E. – Rydén, B.), SNS, Stockholm 1980.

Working Papers

[B7] "What Can Input Tell about Output? – Analyzing Productivity and Efficiency in the Absence of Output Measures" (together with E. Mellander), Working Paper 1987:3, Dep. of Economics, Uppsala University.

[B8] "On the Econometric Analysis of Production, When There Are No Output Data" (together with E. Mellander), Working Paper 1989:6, Dep. of Economics, Uppsala University.

C Econometric Macro Models

Books

[C1] *Two Models of an Open Economy* (ed.), IUI, Stockholm 1986.

[C2] *Kalkyler för 80-talet* (Modeling the 80s) (together with T. Nordström et al.), IUI, Stockholm 1979.

Articles

[C3] Structural Change as an Equilibrium or Disequilibrium Process – An Introduction", in [C1].

[C4] "ISAC – A Model of Stabilization and Structural Change in a Small Open Economy" (together with L. Jansson and T. Nordström), in [C1].

[C5] "Utvecklingsvägar för svensk ekonomi 1978–85 – En kalkylredovisning" (Alternative development paths for the Swedish economy 1978–85 – documenting the projections) (together with L. Jansson and T. Nordström), in [C2] above.

PUBLIC BUDGETING AND FINANCE

D Public Budgeting and Planning

Books

[D1] *Resursfördelning i offentlig budget* (Resource allocation in public budgets), IUI, Stockholm 1982.

Articles

[D2] "Problem of co-ordination of long–range sectoral planning" (together with I. Ståhl), in *Long-term Planning*, U N Sales No E 71 II E 3, New York 1971.

[D3] "Budgetary Inflexibility" (together with A. Robinson), *Journal of Public Budgeting and Finance*, Vol. 2, No. 3, 1982.

[D4] "Re-establishing Flexibility" (together with A. Robinson), *Journal of Public Budgeting and Finance*, Vol. 2, No. 3, 1982.

[D5] "Public Budgets in Sweden", *Journal of Public Budgeting and Finance*, Vol. 2, No. 3, 1982.

[D6] "Varför behövs en budgetreform?" (Why do we need a budgetary reform?), *Tiden*, No. 3, 1965.

[D7] *Civil kontra militär bemanning av statsisbrytarna. En samhällsekonomisk kostnadskalkyl* (Civil versus military staffing of icebreakers – a cost benefit analysis), Ch. 5 in Försvarsdepartementets betänkande 1970:9.

[D8] *Värnpliktskostnaderna* (The costs of general conscription), bilaga till huvudbetänkandet för utredning om handräckningsvärnpliktiga, 1972.

[D9] "Spelet om statens resurser" (The public resource allocation game), *Ekonomisk Revy*, 1973:2.

[D10] "Politik och marknad i offentlig sektor" (Politics and markets in the public sector), in *Politik och Marknad* (ed. Södersten, B.), Lund 1987.

Working Papers

[D11] *Systemanalys och långtidsplanering* (The analysis of systems and long-range planning) (together with I. Ståhl), FOA–P report, C 122, 1964.

[D12] *Allmänna principer för planering och stridsekonomisk värdering*, (General principles of military planning and the economic evaluation of battle effort) (together with I. Ståhl), FOA–P report, C–8171, 1967.

E Taxes and Public Finance

Books

[E1] Bolagsskatt och investeringsvilja (Business taxation and investments) (together with J. Södersten), Research Report No. 17, IUI, Stockholm 1983.

Articles

[E2] "Taxes and Market Stability", in *Business Taxation, Finance and Firm Behavior* (eds. Eliasson G. and Södersten, J.), IUI, Stockholm 1981.

[E3] "Varför utgiftsskatt?" (Why an expenditure tax?), *Ekonomisk Debatt*, Vol. 4, No. 7, 1976.

[E4] *Offentlig ekonomi i tillväxt* (The growing public sector), IUI, Stockholm 1979.

[E5] "Våra skatter 1950–2000 – Ett långsiktigt perspektiv på offentliga skatteforskningsuppgifter" (Our taxes 1950–2000 – a long-term perspective on research tasks in public economics), in *Svensk skatteforskning*, Humanistisk–Samhällsvetenskapliga forskningsrådet, 1979.

[E6] *Fem avgiftsargument – Några principiella synpunkter på finansieringsalternativ för den offentliga sektorn* (The pricing of public services – current practice and future alternatives) (reprint from SOU 1979:23), IUI, Stockholm 1979.

[E7] "I stället för skatterevolt" (Instead of a taxpayer's revolt), in *Svensk ekonomisk politik* (eds. Lundberg, E. and Rydén, B.), SNS, Stockholm 1980.

[E8] *Proportionell eller progressiv vinstskatt?* (Proportional or progressive corporate income taxes?), Ds Fi, 1983:14.

POLICY ANALYSIS

F Methods of Policy Analysis

Articles

[F1] "L'évaluation des Politiques Publiques en Suède", in *L'évaluation des Politiques Publiques* (eds. Nioche, J.-P. and Poinsard, R.), Economica, Paris, 1984.

[F2] "Evaluation of Public Policies", in *IUI Annual Report 1987*, Stockholm.

[F3] "Ekonomiska Rådet – en brygga mellan forskning och förvaltning" (The Economic Council – a bridge between research and public management), in *Konjunktur och Prognos*, Ekonomiska Rådet, 1988.

G Structural and Growth Policy

Books

[G1] *Offentlig service och industriell tillväxt – Perspektivskisser av svensk ekonomisk utveckling 1950–2000* (Public service and industrial growth – Swedish economic perspectives 1950–2000) (together with T. Nordström), Research Report No. 11, IUI, Stockholm 1980.

Articles

[G2] "Strukturpolitik och samhällsplanering" (Industrial policy and economic planning) (together with I. Ståhl), *Tiden*, No. 7, 1965.

[G3] "Strukturinstitutet – en idéskiss (The industrial policy institute – an idea) (together with I. Ståhl), *Tiden*, No. 8, 1965.

[G4] "Strukturproblem och rationaliseringsuppgifter på 80-talet" (Structural problems and the rationalization tasks of the 80s), in *För statlig hushållning under 300 år*, Statskontoret, 1981.

H Local Government Policy

Books

[H1] *Control of Local Governments* (coed. with E. Gramlich), IUI, Stockholm 1985.

[H2] *Kontrollen av kommunerna* (Control of local governments) (together with R. Murray), IUI, Stockholm 1983.

Articles

[H3] "Local Government and Economic Growth", in *The Firm in the Market Economy*, IUI, Stockholm 1980.

[H4] "Accountability, Efficiency and Flexibility in Local Government Behavior", in [H1].

[H5] "Local Authorities, Economic Stability and the Efficiency of Fiscal Policy" (together with T. Nordström), in [H1].

[H6] "Trend och timing i kommunernas utveckling" (Trends and timing in local government development), *Kommunalekonomisk Tidskrift*, No. 3, 1983.

[H7] "Kommunerna i stabiliseringspolitiken" (Local governments and stabilization policies), in *Skuldfällan*, (Bentzel, R. et al.), SNS, Stockholm 1983.

I Industrial Policy

Articles

[I1] "Problems of Maintaining Efficiency under Political Pressure" (together with G. Eliasson), in *State Investment Companies in Western Europe* (ed. Hindley, B.), MacMillan, London 1983.

[I2] "Teori för börskrasch" (A theory of a stock market collapse), *Ekonomisk Debatt*, Vol. 4, No. 4, 1976.

[I3] "Varför löntagarfonder?" (Why wage earners' funds?), in *Sju socialdemokrater om löntagarfonderna* (ed. Lundgren, N.), Tiden, 1979.

[I4] "Postscriptum till fondfrågan" (Postscript on the wage earners' funds issue), in *Sju socialdemokrater om löntagarfonderna* (ed. Lundgren, N.), Tiden 1979.

J Energy Policy

Books

[J1] *Policy Making in a Disorderly World Economy* (coed. with Eliasson, G. and Sharefkin, M.), IUI, Stockholm 1983.

[J2] *Energy and Economic Adjustment* (ed.), IUI, Stockholm 1983.

[J3] *Energy in Swedish Manufacturing* (ed.), IUI, Stockholm 1983.

Articles

[J4] "Oil Prices and Economic Stability – The Macroeconomic Impact of Oil Price Shocks on the Swedish Economy", *Energy*, Vol. 8, No. 5, 1983. A modified version of this paper was also printed in *Adapting to Changing Energy Prices* (eds. Waverman, L. and Watkins, C.), Cambridge, Mass. 1984.

[J5] "Oil Prices and Economic Stability – Simulation Experiments with a Macroeconomic Model" (together with T. Nordström), in [J2] above.

[J6] "Oil Price Instability and Economic Flexibility", in [J2] above.

[J7] "Measuring Energy Substitution", in [J3] above.

[J8] "Energy in Swedish Manufacturing 1980–2000" (together with T. Nordström), in [J3] above.

[J9] "Oljeförbrukningen under 80–talet" (Oil use in the 80s), in *Industri, politik och samhällsekonomi*, IUI, Stockholm 1981.

Working Papers

[J10] *Energi, stabilitet och tillväxt i svensk ekonomi* (Energy, stability and growth of the Swedish economy), Working Paper No. 36, IUI, Stockholm 1981.

K Labor Market Policy

Books

[K1] *Lönebildning och samhällsekonomisk stabilisering* (Wage formation and economic stability) (together with L. Calmfors, B. Holmlund, A. Lindbeck and N. Lundgren), Svenska Sparbanksföreningen, 1985.

Articles

[K2] "Relief Work and Grant Displacement in Sweden", (together with E.M. Gramlich) in *Studies in Labor Market Behavior: Sweden and the United States* (eds. Eliasson, G., Holmlund, B. and Stafford, F.), IUI, Stockholm 1981.

[K3] "Arbetsmarknadsförsäkring – ett reformutkast (Unemployment insurance – a sketch of a policy reform), *Tiden*, No. 3, 1967.

[K4] *Mål och medel i arbetsmarknadspolitisk organisation* (Ends and means in labor market policy), Ch. 10 in SOU 1968:62.

[K5] "Budgetens regionala effekter" (The regional effects of public budgets) (together with P. Lagerblad), in the anthology *Förvaltningsekonomiska problem* (Problems in public management) (ed.), JSF, 1972.

[K6] "Den offentliga sektorn och sysselsättningspolitiken" (The public sector and employment policy), *Ekonomisk Debatt*, Vol. 2, No. 7, 1974.

L Stabilization and Income Policy

Books

[L1] *Blandekonomi i kris?* (The mixed economy in crisis?) (together with S. Grassman, E. Lundberg and I. Ståhl), SNS, Stockholm 1978.

[L2] *Att välja 80-tal – IUIs långtidsbedömning 1979* (Choosing the 80s – IUIs medium-term survey 1979) (ed. together with G. Eliasson and B. Carlsson).

[L3] *Skuldfällan* (The debt trap) (together with R. Bentzel, V. Bergström and I. Persson-Tanimura, I.), SNS, Stockholm 1983.

Articles

[L4] *The Rigors of Welfare – Social Security and Economic Adjustment in Sweden*, Urban Welfare Research Institute, Tokyo, 1981.

[L5] "Income Formation in a Mixed Economy", in *Sweden, Choices for Economic and Social Policy in the 1980s* (eds. B. Rydén and V. Bergström), London 1982.

[L6] "Kalkyler för 80–talet" (Modeling the 80s), Ch. 6 in [L2] above.

[L7] "Prisstabilisering kräver inkomstpolitik" (Price stabilization requires an incomes policy), *Ekonomisk Debatt*, No. 2, 1983.

[L8] "Skattebaserad inkomstpolitik – principer och problem" (A tax based incomes policy – principles and problems) (together with B. Holmlund), *Ekonomisk Debatt*, No. 6, 1983.

[L9] "Stability and Macroeconomic Policy – The Lesson of the 1970s" (together with G. Eliasson and M. Sharefkin), in [J1] above.

[L10] "Konjunkturprognoser: När, Var, Hur?" (Business cycle forecasts – when, where and how?), in *Konjunktur och Prognos* (together with T. Persson), Ekonomiska Rådet, 1988.

M Educational Policy

Books

[M1] *Learning and Earning* (together with M.J. Bowman and Å. Sohlman), Liber, Stockholm 1978.

Articles

[M2] "Homogeneity in Education", in [M1] above.

[M3] "Samordnad samhällsutbildning" (Coordinated education in social sciences), *Tiden*, No. 3, 1965.

M4] *Specialhögskolor för samhällstekniker* (Special universities for social sciences), report from Socialhögskolan in Umeå 1966:1 (together with L. Sköld).

[M5] "Ekonomiska aspekter på läromedelssystem" (Economic aspects on textbook production), *Nordisk utredningsserie*, 1969:16.

[M6] "Effektivitet i högskoleorganisation" (Efficiency and university organization), Ch. 3.2 in *Utvärdering av en högskolereform*, UKÄ, report 1973:2.

[M7] "Allokering genom utbildning" (Allocation through education), Ch. 3.10 in *Utvärdering av en högskolereform*, UKÄ, report 1973:2.

[M8] "Varför högre utbildning?" (Why higher learning?), *Ekonomisk Debatt*, Vol. 1, No. 6, 1973.

[M9] "Den framtida högskolan – ett organisationsalternativ" (The future university – an alternative organization), *Ekonomisk Debatt*, Vol. 2, No. 3, 1974.

[M10] "Högre utbildning och arbetsmarknad" (Higher learning and the labor market), *Ekonomisk Debatt*, Vol. 17, No. 4, 1989.

N Defense Policy

Books

[N1] *Säkerhetspolitik och försvarsplanering* (The policy of security and defense planning) (together with L. Grape), SNS, Stockholm 1967.

Articles

[N2] *International Limitations of Military Expenditure: Issues and Problems* (together with A. Becker), Rand, R–1911–ACDA Santa Monica, 1976. (Later reprinted in part in Becker, A.: Military Expenditure Limitation for Arms Control – Problems and Prospects, Rand, 1977.)

[N3] "En programbudget för försvaret" (Program budgeting in defense), *Balans*, No. 5, 1962.

[N4] "Spelet om världsfreden" (The game of world peace), *Tiden*, No. 3, 1963.

[N5] "Ekonomiska utgångspunkter och överväganden" (Economic premises and considerations in defense policy), Ch. 6 in *Säkerhetspolitik och försvarsutgifter*, SOU 1968:10.

[N6] "Erfarenheter av försvarets långtidsplanering" (Experience from long-range defense planning), in the anthology *Förvaltningsekonomiska problem* (Problems in public management) (ed.), JSF, 1972.

Working Papers

[N7] *Arms Control and Budget Planning* (together with I. Ståhl and E. Moberg). Paper presented at the second Nordic conference on peace research, 1966.

[N8] *Långtidsplanering av vårt lands försvar* (Long-range planning of Swedish defense) (together with Fehrm, Jennergren, Grape and Ståhl), FOA–P rapport A 143, 1965.

O Housing Policy

Articles

[O1] "Det stora bostadslotteriet – bostadspolitikens fördelningseffekter" (The housing lottery – the distributional effects of housing policies), *Ekonomisk Debatt*, October 1974.

[O2] "Skatteprogressivitet och småhuspriserna" (Tax progressivity and home prices), *Ekonomisk Debatt*, Vol. 4, No. 5, 1976.

P Other Policies

Articles

[P1] "Hur värderas människoliv" (The economic value of human life) (together with I. Ståhl), *Läkartidningen*, No. 39, 1965.

Publications in English

1991

Truth and Meaning in Economics – Selected Essays on Economic Theory and Policy. Bengt-Christer Ysander. 193 pp.

The Rate of Return and the Rate of Interest. Robert M. Solow. 36 pp.

1990

The Knowledge Based Information Economy. G. Eliasson, S. Fölster, T. Lindberg, T. Pousette and E. Taymaz. 182 pp.

International Financial Integration. Lars Oxelheim. IUI, Springer-Verlag. 389 pp.

Growth and Integration in a Nordic Perspective. IUI, ETLA, IFF, NØI. 225 pp.

1989

MOSES Code. James W. Albrecht et al. Research Report No. 36. 354 pp.

MOSES Handbook. Fredrik Bergholm. Research Report No. 35. 213 pp.

1988

Factor Demand in Swedish Manufacturing: Econometric Analyses. Joyce Dargay. Research Report No. 34. 138 pp.

The Knowledge Base of an Industrial Economy. Gunnar Eliasson. Research Report No. 33. 100 pp.

The Political Economy of the Welfare State. James Buchanan. 31 pp.

1987

Technological Competition and Trade in the Experimentally Organized Economy. Gunnar Eliasson. Research Report No. 32. 118 pp.

The European Economic Recovery – A Need for New Policies? Franco Modigliani. 46 pp.

Analyses of industrial Structure – A Putty-Clay Approach. Finn R. Førsund and Lennart Hjalmarsson. 321 pp.