ESSAYS IN COMPARATIVE DYNAMICS

by

RUSSELL DAVIDSON

B.Sc., University of Glasgow, 1963,
Ph.D., University of Glasgow, 1966.

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
(Department of Economics)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
July, 1977

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of ECONOMICS

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date 8th AUGUST 1977.

Research Supervisor: Professor G.C. Archibald.

ABSTRACT.

Problems in the theory of economic dynamics are tackled both by theoretical arguments and by use of specific examples. The work is divided into three essays. The first treats optimal control theory from an economic point of view, giving an exposition of the mathematical theory in terms of economic concepts. The idea of the marginal worth of time is introduced and found to be useful in a variety of problems. An interpretation is given of the phase planes of optimal control problems as demand-and-supply diagrams. The second essay makes use of the techniques developed in the first to solve the problem of when and how a firm faced with adverse economic circumstances will choose to go out of business if its operations depend on a stock of some fixed asset that depreciates over time. A straightforward catalogue is presented of different possible outcomes. The third essay deals with a model of urban housing. It contains two main sections. In the first, an equilibrium state is described in which demand by tenants for housing is met by supply from landlords who act as profit maximisers over the whole period of time that their property exists. The rent paid for any particular dwelling is assumed to depend on its state of upkeep, which in turn depends on how much is spent by a landlord on maintenance. The equilibrium is found by a procedure analogous to that regularly used in general equilibrium theory, namely by finding a fixed point of a mapping in a (here infinite-dimensional)

vector space. In the second section of the essay, it is assumed that some externality arises which adversely affects urban life and which provokes people to move out to suburbs. The consequences of this are studied and two different kinds of dynamical evolution can be distinguished. One, in which house construction in the suburbs is slow enough to make it necessary for new construction to continue in the city, tends not to be disastrous for the city; the other, in which all urban construction stops when the externality arises, usually leads to complete decay of the city. Throughout the thesis there is an emphasis on the differences in approach between static or quasistatic problems and dynamic ones.

CONTENTS.

LIST OF TABLES.

LIST OF FIGURES.

ACKNOWLEDGEMENTS.

It is with much pleasure that I acknowledge the great deal of help and support that I have received from many members of the UBC Economics Department. First, my heartfelt thanks go to Don Paterson, who rescued me from Ec. 100 and persuaded me to study economics more seriously. When I joined the department, I was encouraged and stimulated by Chuck Blackorby, David Donaldson, Curt Eaton, Keizo Nagatani, Phil Neher and many others. The debt I owe them is immense. Special thanks must go to Chris Archibald, chairman of my thesis committee, who, no less an intellectual companion than anyone in the department, managed also to make me comply with deadlines and the requirements of procedure, things I am not very good at. It is the friendship offered me by all the people I have named that I rejoice in most, and what I say here must be quite inadequate.

Grateful acknowledgement is made of financial support from the Canada Council, in the form of a Doctoral Fellowship.

INTRODUCTION.

The common theme of the three essays of this thesis is, as the title would suggest, economic dynamics. A desire to see into the future, with especial regard for one's own fortunes and well-being, has been implanted in mankind from the beginning. Necromancy and augury are among the time-honoured techniques for accomplishing this desire - they have a much longer history than scientific, or even pseudo-scientific, analysis, and numerology and astrology are their spiritual children. The study of economics is often grouped with the above practices, not necessarily to the distress of its professional advocates, but very much to the denigration of its avowed means and ends. Imputations of wizardry can be hard to deny although they can be ignored by practitioners of mature sciences. Ignoring is perhaps the wrong response, for it provides no corrective to superstition when needed. Thus the natural philosophers following Galileo and Newton were content that they had escaped from the mental chains of medieval thought and alchemy, but did not always understand that their achievements were taken by many as just somewhat *better* magic than what was fashionable in earlier times. So it was that their spells were seized upon by lesser men as panaceas for all earthly disorders, and even by some great men struggling with matters less tractable than the inanimate physical universe.

For Marx had his incantation quite correct. The "laws of motion of capitalism" is a phrase that has called up a spirit by no means exorcised

in our times. The great difficulty with such spirits is that no one knows if they can ultimately be taught the catechism and made to sit with all proper appurtenances and apparel in the halls of true science and knowledge, or if they are of the other kind of spirit, demons, which do nothing but torment our minds with vain longings that can never be satisfied.

I have sported with this spirit, then, in these essays. My leaning is towards the hall of science, and my effort is directed towards developing a little piece of economic theory that will help the understanding of what sorts of questions about economic dynamics can reasonably be treated by use of the wits and intuition that the human race has succeeded in getting for itself at the present time.

Discovering the most sensible questions to investigate is one of the difficult things about the study of comparative dynamics. It is for this reason that I do not start off, in decent fashion, by defining what I mean or what is generally meant by comparative dynamics. Comparative statics is nowadays an established discipline, its rules laid out as formulae to be applied mechanically. In a sense, it would be good if comparative dynamics were in the same case, so that there is in this thesis a good deal of laying down of rules and procedures. But to define is to limit, and since no one knows where it might be best to set the limits of comparative dynamics, it is foolish to define with any precision. An alternative way to go forward is to look at specific problems that involve dynamical considerations, and that is the way chosen here. It points up the conclusion that comparative dynamics has a quite different flavour from comparative statics. It is not just that the well-known techniques of total differentiation followed

by attempts at signing bordered Hessians no longer seem so immediately applicable, but also that the economic matters that are to be taken into account have no analogue in static problems - they are not just extensions of static concepts with time thrown in as an extra variable.

The tone of the three essays that comprise the thesis varies somewhat from one to another. This is due to the variety of tasks undertaken. In the first essay, the aim is to consolidate as a standard box-of-tricks for economists' use the mathematical theory of optimal control. There are numerous economic studies that make use of this theory, often very expertly. Consequently my emphasis has been on expounding, on demystifying as it were, and the resulting tone is allusive and sometimes chatty, my hope being that it is also evocative and illuminating. The second essay arose from attempts to solve two seemingly unconnected problems. The first one, which in my mind has attached to it the name of "the bankrupt railroad", was suggested to me by Professor Archibald and deals with the response of a firm which possesses substantial fixed assets when it runs into difficulties. The second problem was posed to the Economics Department at large by Professor Nagatani and is, essentially, the one explicitly treated here. The two problems are the same on a suitable level of abstraction, both contain a basic and typical question in economic dynamics, and both are solved by the same device, the explication of which forms the content of the essay. The use of mathematics in the last essay of the thesis is much heavier than in the other two, and the effect of this is - to my regret - to make the tone much heavier too. But it is unavoidable at present - instances are very rare of a new sort of investigation being presented for the first time in its most perspicuous form. I hope that the gain in understanding is worth the cost.

Optimal control theory is what makes comparative dynamics possible, at least in a practical sense. F.P. Ramsey's pioneering article (Ramsey (1928)) makes no use of it because it wasn't invented then, but it is quite fascinating for someone who knows the modern theory to see how much of it appears there in only a slightly different form. His formal tool is of course the calculus of variations, and it can now be seen that, if one works hard enough at it, (Hadley and Kemp (1971)) optimal control theory can be *derived* from the calculus of variations. But the insights gained by physical scientists as well as economists from the twin notions of the principle of optimality (Bellman (1961)) and the maximum principle (Pontryagin et al (1962)) are much more numerous than those that the classical approach can provide. It is striking that Ramsey supplements his calculus of variations argument with an economic line of reasoning that he ascribes to Keynes and that has much of the spirit of the two principles which underlie optimal control theory. The first essay, then, is in the tradition of Ramsey-cum-Keynes and Bellman, and away from the work of Hadley and Kemp.

Growth models have been at the centre of economic dynamics at least since Ricardo, and certainly with Marx, although it was not exactly growth in the modern sense that Marx was concerned with. Modern interest in growth theory began with Harrod (1939) and Domar (1946), but no one at that time conceived of economic dynamics as other than a descriptive discipline that could, to be sure, warn against some kinds of danger. (Ramsey's article stands outside this discussion - it was really much ahead of its time.) Rates of capital accumulation were derived from purely *ad hoc*

descriptions of saving behaviour, and the social benefits arising from different specifications of behaviour compared. It would be wrong to deny the name of comparative dynamics to these exercises, but there was precious little economising going on in the models used. That is, there were no economic agents exercising rational choice over their behaviour. This was true also of the celebrated von Neumann growth model (von Neumann (1945)), and even of the subtle model proposed by Solow (1956) with its exogenous savings ratio. Phelps (1961), in his critique of Solow's model, returned in spirit to Ramsey, and the word "optimal" reappears in the growth literature, just at the time when by a fortunate coincidence U.S. and U.S.S.R. military needs caused Bellman and Pontryagin respectively to produce what is now called optimal control theory.

By the time of writing of articles by Kurz (1968) and Hahn (1968), this theory was taken for granted by growthmen, and all sorts of problems to do with stability and so forth were being encountered. These problems have led sophisticated mathematical economists like Brock (1976), Brock and Burmeister (1976), Brock and Scheinkman (1975) (see also other articles listed by these authors in the cited works and the "Symposium on Hamiltonian Dynamics in Economics" published as the Feb. 1976 issue of the *Journal of Economic Theory*) to study intently the stability of dynamical systems and to propose the resurrection of Samuelson's suggestion that the assumption of stability can yield meaningful theorems in economics. (I cannot resist citing Hatta (1977) as a beautiful example of an article that uses this suggestion with great effect in comparative statics.)

Of more immediate relevance, it seems to me, than this very technical work is the systematic use of optimal control theory in microeconomic studies as opposed to the macroeconomics of growth. The principal example of this that I have in mind is Mortensen's article on wage and employment dynamics (Mortensen (1970)), which has been followed by various articles on job search by employees and labour hiring by employers written from an explicitly dynamical point of view. In this literature the optimising behaviour of economic agents comes to the fore. It is certainly premature to claim that economic theory contains a worthy set of "micro-foundations" for macroeconomic phenomena, but it is hard to doubt that such foundations will be necessary for a decent understanding of them. It is my hope, then, that applications of optimal control theory will ultimately provide enough insight into comparative dynamics in microeconomics that one will be able to return to growth theory and actually succeed in predicting a little bit of the future of our fortunes.

CHAPTER I

AN ECONOMIC INTERPRETATION OF OPTIMAL CONTROL THEORY

In this essay, optimal control theory, the cornerstone of comparative dynamics in its present state of development, is discussed from an economic point of view. Much of the subject matter will seem routine to those who regularly use optimal control methods, but although there are many economists among these people, I have the feeling that most of them share my view that the theory has not yet been fully "civilised", that is, translated out of the language of engineering physics, the discipline responsible for its creation in its modern form, into a language of prices, margins, demands and so on easily comprehensible to economists and such as to make the optimal control equations satisfying to economic intuition. A stylised problem is treated at full length in the remainder of this essay, and it is my hope that the treatment, which is throughout presented in economic language, will act as a sort of translation of the theory.

In section 1, the well-known optimal control algorithm is derived from first principles. A form of argument much used in duality theory is found useful in deriving the equation of motion for the co-state variable. Then, in section 2, comes the main effort of "translation". The algorithm is picked apart and all of its variables, functions, and equations discussed as economic entities. After this work, it becomes possible in section 3 to discuss the matter of transversality conditions as a series of stories in economics, by use of no more than the everyday arguments of economic reasoning. In section 4, I have discussed the notion of time as a factor of production

and the related concept of its marginal worth. It turns out to be possible to define this concept quite precisely, and to use it effectively in the treatment of autonomous problems and problems with discounting -- these problems are worked out as examples. Finally, in section 5, it is shown that, in many cases, the phase-plane, with its saddlepoint equilibrium and catenary-like motion in the vicinity, can be seen as a straightforward supply-and-demand diagram with adjustment mechanisms around equilibrium included: just what is required by Samuelson's correspondence principle in fact (Samuelson [1947], p. 253 et seq., p. 350).

I have tried to keep all the discussion of this chapter uncomplicated as possible as regards the mathematical equipment used. Much greater generality is easily accessible with little extra effort, but it was not at all my aim to indicate the richness of optimal control theory or its applications -- that is done in (probably) hundreds of easily available references. (For example, Bryson and Ho [1969], Bellman [1967], Lee and Markus [1967]). My aim, on the contrary, was to demonstrate that the theory could just as well have been developed by economists, using their own tools, techniques and language, as by engineers, if only their research needs in the late fifties and early sixties had been so clamant and their work so heavily funded.

Such a demonstration could well be quite pointless, but it does seem to me that the concepts developed here are both useful, and in some cases, novel. So far as I know, no one, not even Arrow [1968], has yet pointed out in the economics literature that the marginal worth of time is constant along optimal paths; and the interpretation of the phase-plane as a supply-and-demand diagram is certainly new.

1. Dynamic Programming and Optimal Control

The problem to be considered in this section is that of maximising the functional

$$J[c] \equiv \int_{t_0}^{t_f} F(k,c,t)\ dt \tag{1}$$

where c, the argument of J, is a function of time t, defined on the interval $t_0 \leq t \leq t_f$, and the function k, defined on the same interval, is given as the solution of the ordinary differential equation

$$\dot{k}(\equiv \frac{dk}{dt}) = f(k,c,t) \tag{2}$$

with boundary condition

$$k(t_0) = \bar{k}_0 \tag{3a}$$

We restrict the domain of functions c over which maximisation takes place by imposing another boundary condition

$$k(t_f) = \bar{k}_f \tag{3b}$$

The problem is *controllable* if there exist functions c which allow equation (3b) to be satisfied. Thus k is completely specified once c is given. In the expressions F(k,c,t) and f(k,c,t), what is meant is that both F and f are functions of three variables, and that each is evaluated at the point (k(t), c(t), t). But for simplicity the arguments of k and c will be omitted unless confusion is likely.

The variable names, k and c, are meant to be suggestive. One may think of k as a capital stock and of c as a rate of consumption. Then the differential equation (2) will express the rate of capital accumulation (investment) as a function of the existing capital stock, k (let the labour supply be fixed, for example), and the rate at which output is

consumed. A common form that equation (2) might take is

$$\dot{k} = g(k) - \delta k - c$$

where g is a production function and δ a depreciation rate. The equation tells us that consumption plus net investment, $c + \dot{k}$, equals total output g(k) less the amount of output, δk, needed to offset depreciation. The function F(k,c,t) may perhaps measure the utility of consumption, or the profitability of producing consumption goods, or in general the worth of some benefits. The fact that J[c] is given as an *integral* means that these benefits accrue *additively* over time. If F is interpreted as a utility function, then this form of J[c] implies additive utility in the intertemporal sense; that is, no intertemporal complementarity. Since this notion is artificial, it is probably better to construe F as a measure of profit. I shall try to refer to it consistently as the "benefit".

The equations (1) and (2) define a standard optimal problem, and it is as well to introduce the standard terminology now. The function c, which determines, both directly via its appearance in the benefit function F, and indirectly via k, the value of "total benefit" J, is called the *control variable*. We can assume that c can be chosen quite arbitrarily, or, if it is preferable, we can restrict it in some way. The two cases are equally easy to handle, conceptually at least. It is important to understand that the control variable c, and *only* c, is at our disposal. Within the confines of whatever restrictions are imposed, it is to be chosen so as to maximise benefits, that is, J. The other variable, k, is

not at our disposal, except indirectly. Once c is chosen, k is *determined* by equation (2) and one of the boundary conditions (3). Consequently, k is called the *state variable*. It affects benefits (by appearing as an argument in F) and so has to do with the *state* of affairs.

In order to find the controller (as it is often called), c*, which gives a maximum of J[c], we shall use the approach of dynamic programming. This approach rests on a very general principle, the *Principle of Optimality* (see Bellman [1961]). According to this, whenever there exists an optimal way of achieving some end or of carrying out some activity which proceeds by a sequence of steps, each step contributing additively to a "performance criterion" (for our purposes, the criterion is J[c]), then if one breaks in on the sequence part of the way through, the steps from that point of break-in until the end of the sequence must be optimal for the problem defined over those stages alone. For our problem, this means that if c* is the optimal controller for the problem of equations (1) and (2) with boundary conditions (3), then the same function c*, is the optimal controller for the problem:

$$\max \hat{J}[c] \equiv \int_{t_1}^{t_f} F(k,c,t)\, dt$$

with $\dot{k} = f(k,c,t)$ and boundary conditions (4)

$$k(t_1) = k^*(t_1)$$

$$k(t_f) = \bar{k}_f$$

where by $k^*(t_1)$ we mean the value of k arrived at by time t_1 if the path defined by c* has been followed from time t_0 with $k(t_0) = \bar{k}_0$.

The Principle of Optimality is both general and trivial. Its truth follows from a short *reductio ad absurdum* argument: If another function, c**, say, defined for $t_1 \leq t \leq t_f$ gave a larger value of J[c] than c*, then the function defined by

$$c^+(t) = \begin{cases} c^*(t) & t_0 \leq t \leq t_1 \\ c^{**}(t) & t_1 \leq t \leq t_f \end{cases}$$

would yield a larger value of J[c] over $t_0 \leq t \leq t_f$ than would c*. But this contradicts the definition of c* as the optimal controller. This kind of argument is precisely the one used in all stepwise optimisation procedures.

Back to the problem at hand. Let us define another functional closely related to J[c]:

$$J(k_t, t, \underset{t \leq \tau \leq t_f}{c(\tau)}) \equiv \int_t^{t_f} F(k(\tau), c(\tau), \tau)\, d\tau$$

where k is defined by

$$\dot{k} = f(k,c,t)$$

$$k(t) = k_t; \quad k(t_f) = \bar{k}_f.$$

This then is a functional of the function c, defined on $t \leq \tau \leq t_f$, with extra dependence on the numbers k_t and t. This new J is just the performance criterion to be maximised for a problem like the old one in all aspects except its time horizon and its initial value of the state variable. Now let us imagine that the maximisation has been performed. Then we will write:

$$J^*(k_t,t) = \max_{c(\tau)} J(k_t, t, \underset{t \leq \tau \leq t_f}{c(\tau)})$$

The star indicates that the maximisation has been done. There will be a certain function, c*, say, which causes J to take on its optimum value J*. That is:

$$J^*(k_t,t) = J(k_t,\ t,\ \underset{t\leq\tau\leq t_f}{c^*(\tau)})$$

Now

$$J(k_t,\ t,\ \underset{t\leq\tau\leq t_f}{c^*(\tau)}) = \int_t^{t_f} F(k^*(\tau),\ c^*(\tau),\ \tau)d\tau$$

[k* is the solution of the defining differential equation with c = c*].

The integral can be split up:

$$J(k_t,\ t,\ \underset{t\leq\tau\leq t_f}{c^*(\tau)}) = \int_t^{t+\delta t} F\ d\tau + \int_{t+\delta t}^{t_f} F\ d\tau$$

Now for the Principle of Optimality: the second term here $\int_{t+\delta t}^{t_f} F\ dt$ must be the *optimised* value for the problem beginning at t+δt with boundary condition k(t+δt) = k*(t+δt). This means in fact:

$$\begin{aligned} J(k_t,\ t,\ \underset{t\leq\tau\leq t_f}{c^*(\tau)}) &= J^*(k_t,t) \\ &= \int_t^{t+\delta t} F(k^*(\tau),\ c^*(\tau),\ \tau)d\tau \\ &\quad + J^*(k^*(t+\delta t),\ t+\delta t). \end{aligned} \tag{5}$$

It should be noticed here that J* is simply an ordinary *function* of the two variables k_t and t, once the maximisation has been done. If we may cheerfully assume all the continuity and differentiability that we need,

then J^* can be expanded in a Taylor series about (k_t, t). (It is not the intention here to worry about technical details, and such "nice" properties will always be taken for granted. There are entirely reasonable sufficient conditions for these properties to hold. For statement and proof of Taylor's theorem, see Hardy (1952), p. 286.) Thus we may write:

$$J^*(k^*(t+\delta t),\ t+\delta t)$$

$$= J^*(k^*(t),\ t) + J_1^*(k^*(t),\ t)\delta k + J_2^*(k^*(t),\ t)\delta t + o(\delta t).$$

The notation in this equation is as follows: J_1^* and J_2^* are the two partial derivatives of J^*;

$$\delta k \equiv k^*(t+\delta t) - k^*(t)$$

$$= k^*(t+\delta t) - k_t \text{ by the boundary condition;}$$

the expression $o(\delta t)$ denotes any quantity, X, say, such that $\lim_{\delta t \to 0} \frac{X}{\delta t} = 0$. In the equation above, $o(\delta t)$ signifies terms proportional to higher powers of δt than the first (for more details, see Hardy [1952], p. 183).

It follows at once from the differential equation (2), which the function k^* must satisfy that:

$$\delta k = \frac{dk^*}{dt}\delta t + o(\delta t)$$

$$= f(k_t,\ c^*(t),\ t)\delta t + o(\delta t).$$

Therefore equation (5) becomes:

$$J^*(k_t,\ t) = \int_t^{t+\delta t} F(k^*(\tau),\ c^*(\tau),\ \tau)d\tau$$

$$+ J^*(k_t,\ t) + [J_1^*(k_t,t)\ f(k_t,\ c^*(t),\ t)$$

$$+ J_2^*(k\ ,\ t)]\delta t + o(\delta t).$$

The integral of course can be written:

$$\int_t^{t+\delta t} F(k^*(\tau),\ c^*(\tau),\ \tau)\,d\tau = F(k_t,\ c^*(t)\ t)\,\delta t + o(\delta t);$$

so that

$$J^*(k_t,\ t) = [F(k_t, c^*(t),\ t) + J_1^*(k_t,t)\ f(k_t,\ c^*(t),\ t)$$

$$+ J_2^*(k_t,t)]\,\delta t + J^*(k_t,t) + o(\delta t)$$

There are now only two terms which involve the control variable explicitly, and in these only its value at time t appears. The optimality for other times is built into the definition of J*. For time t, it can be made quite explicit:

$$J^*(k_t,t) = J^*(k_t,t) + J_2^*(k_t,t)\,\delta t$$

$$+ \max_{c(t)} [F(k_t,c(t),t) + J_1^*(k_t,t)\ f(k_t,c(t),t)]\,\delta t$$

$$+ o(\delta t).$$

This is true because c*(t) is the *optimal* controller. If we let $\delta t \to 0$, we obtain, recalling the definition of $o(\delta t)$:

$$J_2^*(k_t,t) + \max_{c(t)} [F(k_t,c(t),t)$$

$$+ J_1^*(k_t,t)\ f(k_t,c(t),t)] = 0 \qquad (6)$$

where the value of c(t) which yields the maximum is c*(t). This last equation is called the *Hamilton-Jacobi equation*, by analogy with an

equation known for some time in classical mechanics. Because of its newly-discovered use in dynamic programming, it is also called the Hamilton-Jacobi-Bellman equation (see Bellman [1967], chapter 5).

Equation (6) already contains the whole of the optimal control algorithm for our problem. All that remains for us to do is to dig out of it the well-known form of the algorithm. The content of the equation can be expressed particularly simply if we make the following definition:

$$H(k,c,\lambda,t) = F(k,c,t) + \lambda f(k,c,t) \tag{7}$$

This function of the four variables k, c, λ, t, is called the *Hamiltonian* of the problem. In classical mechanics, the Hamiltonian is identified with the *energy* of a system: it is the function which determines the evolution of a mechanical system through time. In economics, the meaning of the Hamiltonian is somewhat similar, but it is susceptible, all the same, to an intuitively appealing and purely economic interpretation.

By use of equation (7), equation (6) can be written:

$$J_2^*(k_t,t) + \max_{c(t)} H(k_t,c(t), J_1^*(k_t,t),t) = 0$$

with the Hamiltonian function evaluated at $\lambda = J_1^*(k_t,t)$. From the definition of $c^*(t)$, we have:

$$\max_{c(t)} H(k_t,c(t), J_1^*(k_t,t),t) = H(k_t,c^*(t), J_1^*(k_t,t),t)$$

This is the first part of the optimal control algorithm: the optimal controller c^* maximises, *at each point in time*, the value of the Hamiltonian. The Hamiltonian as we are using it, however, depends on $J_1^*(k_t,t)$. But let

us simply treat this quantity as an unknown (as yet) function of time $\lambda(t)$, and then pin down its behaviour by a differential equation. To do this, it is necessary to recall that $k_t = k^*(t)$. The function $\lambda(t)$ is then defined:

$$\dot{\lambda}(t) = J_1^*(k^*(t),t), \tag{8}$$

so that

$$\lambda = \frac{d\lambda(t)}{dt} = \frac{dJ_1^*}{dt}(k^*(t),t)$$

$$= J_{11}^* \frac{dk^*}{dt} + J_{12}^* = J_{11}^* f(k^*,c^*,t) + J_{12}^* \tag{9}$$

This last quantity can be obtained directly by differentiating equation (6) with respect to k_t. But perhaps the most illuminating way to proceed, because of the similarity of the technique to one used constantly (see Gorman[1968]) in duality theory, is to note that equation (6) says:

$$\max_c \{J_2^*(k^*(t),t) + F(k^*(t),c,t)$$

$$+ J_1^*(k^*(t),t)\ f(k^*(t),c,t)\} = 0.$$

Therefore, for any $k \neq k^*(t)$, $J_2^*(k,t) + F(k,c^*(t),t) + J_1^*(k,t)\ f(k,c^*(t),t) \leq 0$ because $c^*(t)$ is the optimal controller if $k_t = k^*(t)$, but *not* if $k \neq k^*(t)$. Equality occurs in this expression only if $k = k^*(t)$, which is thus a maximum point of the expression. So then the derivative with respect to k is zero for $k = k^*(t)$:

$$J_{12}^*(k^*(t),t) + F_1(k^*(t),\ c^*(t),t)$$

$$+ J_{11}^*(k^*(t),t)\ f(k^*(t),\ c^*(t),t)$$

$$+ J_1^*(k^*(t),t)\ f_1(k^*(t),\ c^*(t),t) = 0.$$

But from equations (8) and (9),

$$\dot{\lambda} = J_{11}^{*} f(k^*,c^*,t) + J_{12}^{*}$$

$$= -F_1(k^*,c^*,t) - \lambda f_1(k^*,c^*,t)$$

$$= - \frac{\partial H(k^*,c^*,\lambda,t)}{\partial k} \tag{10}$$

And this is indeed the second equation of the optimal control algorithm. We have therefore justified the use of $J_1^{*}(k^*(t),t)$ as $\lambda(t)$, the *co-state variable*.

Now we have all that we need. The equations (2), (10), and (3) give

$$\dot{k} = f(k,c,t) = \frac{\partial H}{\partial \lambda}(k,c,\lambda,t) \tag{11}$$

$$\dot{\lambda} = - \frac{\partial H}{\partial k}(k,c,\lambda,t) \tag{12}$$

$$k(t_0) = \bar{k}_0;\ k(t_f) = \bar{k}_f \tag{13}$$

to be satisfied by the optimal paths of the variables k, c, λ, and further we know (p. 17) that

$$H(k^*,c^*,\lambda^*,t) = \max_{c} H(k^*,c,\lambda^*,t) \tag{14}$$

This is all of the optimal control algorithm. It is enough to determine the optimal path completely, since (14) gives c^* as a function of k^* and λ^* which can be substituted into equations (11) and (12), two first-order differential equations, with two boundary conditions, equation (13), and thus quite determinate.

Sometimes the algorithm is quoted with a Hamiltonian, $H^*(k,\lambda,t)$ which is the result of the maximisation of equation (14). The differential equations are given simply as

$$\dot{k} = \frac{\partial H^*}{\partial \lambda}(k,\lambda,t)$$

$$\dot{\lambda} = -\frac{\partial H^*}{\partial k}(k,\lambda,t) \qquad \text{(boundary conditions as before)}$$

This is no different from the algorithm as we have quoted it, since, for instance,

$$\frac{\partial H^*}{\partial \lambda}(k,\lambda,t) = \frac{\partial H}{\partial \lambda}(k,c^*,\lambda,t) + \frac{\partial H}{\partial c}(k,c^*,\lambda,t)\,\frac{\partial c}{\partial \lambda}$$

but $\frac{\partial H}{\partial c} = 0$ for $c = c^*$ by the maximality relation (14). [When the maximum of equation (14) is not an interior one, but is on the boundary of the admissible set of c's, then the argument must be modified in a way familiar enough if one knows the Kuhn-Tucker equations. The result is unchanged].

2. Discussion of the Algorithm

The derivation of the optimal control algorithm of the preceding section has been rather mathematical, and it was based on the Principle of Optimality, which is a very general principle. It is possible to give much more economic insight into the workings, now that we have seen all the relevant mathematical relations. The whole problem can be thought of as being the finding of an expression for *derived demand*. The economic notion behind derived demand for anything, be it labour, capital or anything else which yields benefit not only (or not at all) directly but also

by being used in some way, is that, if efficiency of use is guaranteed somehow (as, for instance, by competition for resources in a state of perfect information) then the worth of a stock of the thing is given by the value of the *maximum* benefit the stock can yield (that is, if it is used *optimally*) and the derived demand is the worth at the margin. When the problem of equations (1), (2), (3) is posed then, what the answer, $J^*(\bar{k}_0, t_0)$ tells us is the *worth* of a stock k_0 which can be used between the times t_0 and t_f with the constraint, equation (2). If an economic agent is given the choice of having the stock k_0 under those conditions or not, it follows at once that $J^*(\bar{k}_0, t_0)$ is the demand price (assuming that the agent is rational) that he is willing to pay to have the stock. If $\bar{k}_0$ is divisible on the other hand, then $\frac{\partial J^*}{\partial k_0}$ is the demand price of another unit of stock -- the *marginal* worth of the stock. These remarks apply just as well at any time $t(t_0 \leq t \leq t_f)$ to $J^*(k(t),t)$.

All of the above is just an elaborate statement of the *definition* of the maximum benefit a stock can yield, and consequently the derived demand for it. But it enables us to interpret the various parts of the optimal control algorithm. From equation (6), it is seen that $c^*(t)$ is the value of the control variable at time t which maximises

$$F(k(t),c(t)t) + J_1^*(k(t),t)\ f(k(t),c(t),t) \tag{15}$$

Now $F(k,c,t)$ is the integrand of the objective functional J, and so it measures the *rate* at which (at time t) benefits are accruing. We may call it the *current rate of satisfaction*. $J_1^*(k(t),t)$ has just been

identified as the marginal worth of stock -- the derived demand for another unit of it. But $f(k,c,t)$ is just $\dot{k}$, the rate of accumulation of stock. So the second term in the expression (15) measures the rate of accumulation of benefits to be reaped in the future. Thus the whole expression (15) measures the total rate of acquisition of benefits, both current and expected and it is not hard to see why $c^*(t)$ is chosen to maximise it. The optimal $c^*(t)$ gives that particular trade-off between current satisfaction, $F(k,c,t)$, and future satisfaction which maximises total benefit. Since, once $\lambda(t)$ is substituted for J_1^* in expression (15), we have just the Hamiltonian, it follows that the Hamiltonian is the rate of acquisition of benefits -- or, in other words, the benefits that derive from the *duration* of one unit of time. This latter interpretation will be much expanded later.

In the language of economic dynamics, the requirement that $c^*(t)$ should maximise expression (15) is the requirement of *equilibrium at each moment*, let us say of "current equilibrium". That is, it expresses an equilibrium between current needs as expressed through direct demand for current satisfaction and future needs as expressed through the derived demand for stock. $\lambda(t)$, as usual for Lagrangian multipliers in economics, is the equilibrating *price*, if, as has been implicit throughout this discussion, "benefit" is taken as numeraire. (For $\lambda = J_1^*$ = marginal worth (i.e., benefit) of stock). With this in mind, it is not surprising that equations (11) and (12) provide the rest of the economic dynamics, that is the link between successive current equilibria. Equation (11) is our *physical* restraint, given to us exogenously, and can be

thought of, as in the example given in the last section, as a production function. So much is very easy. But now it is clear that equation (12) is a price adjustment mechanism reflecting the changes over time in the marginal valuation of stock.

What is the sense of this price equation? One remark here as a bit of a digression. Now that λ is identified as a (shadow) price, or marginal evaluation of something, it is plain why equation (12) emerged (see p. 17) through an argument borrowed from duality theory. Duality theory reflects various symmetries between quantities and prices, and its techniques let us go from conclusions about one set of these variables [eq. (6)] to conclusions about the other [eq. (12)]. Let us write out equation (12) more explicitly:

$$\dot{\lambda} = -\frac{\partial H}{\partial k} = -F_k(k,c,t) - \lambda f_k(k,c,t),$$

whence

$$\frac{\dot{\lambda}}{\lambda} + f_k(k,c,t) + \frac{1}{\lambda} F_k(k,c,t) = 0 \tag{16}$$

The first term is the rate of accrual of capital gains to holders of stock (λ is its price). $f_k(k,c,t)$ is the own rate of return on stock, since it is the increment to the rate of accumulation $\dot{k}$ arising from one unit more of stock. $F_k(k,c,t)$ is the increment to the current rate of satisfaction from one unit more of stock, and since $\frac{1}{\lambda}$ is the price of satisfaction in terms of stock, the third term in equation (16) is the revenue to a stock holder from providing this increment of satisfaction. Equation (16), then, is a zero net-profit condition, and it consequently confirms the identification of λ with a shadow price. It says that the total rate of return on the

marginal unit of stock is zero: in one unit of time, and for this marginal unit, the increase (in stock units) of its value ($\dot{\lambda}/\lambda$) plus the produced increase in stock (f_k) plus the value (in stock units) of the increased current satisfaction ($\frac{1}{\lambda} F_k$) add up to zero.

A point of possible confusion: What about "normal" profits? Surely the rate of return should equal the rate of interest, not zero? The answer is easy: λ is a price quoted *at time* t_0, since J^* is measured always in the same way. That is, $J^*(k(t),t)$ is defined as

$$\max \int_t^{t_f} F(k(\tau), c(\tau), \tau)\, d\tau$$

which is the value, *as seen from* t_0, of the part of the programme from t to t_f. Consequently, if there is a rate of interest, $\dot{\lambda}/\lambda$ is the rate of capital gains (in the usual sense) *minus* the rate of interest: for example, what one usually calls "no capital gains" corresponds to *present-value* prices declining into the future at the interest rate, i.e., $\frac{\dot{\lambda}}{\lambda} = -r$. A fuller treatment of this matter will be given later.

3. Transversality

The problem considered so far, as defined by equations (1), (2), (3) has been that of maximising the functional $J[c]$ for a fixed range of time t_0 to t_f, and with a requirement that, at time t_f, the stock or state variable k should take on the value $\bar{k}_f$. It is frequently the case in problems in economics that neither t_f nor k_f is specified, but that both can be chosen optimally. There is conceptually no difficulty whatever to this. The function $J^*(\bar{k}_0, t_0)$ is worked out for a range of accessible values of t_f

and $\bar{k}_f$, all that is needed is to find out the point where the appropriate first derivatives vanish. It is convenient now to include as arguments of J* the terminal quantities t_f and k_f (dropping the bar for clarity) and to suppress $\bar{k}_0$ and t_0, it being understood that these are given and fixed. Thus we define:

$$J(t_f,k_f) = \int_{t_0}^{t_f} F(k^*,c^*,\tau)\ d\tau$$

where $k^*(\tau,t_f,k_f)$ and $c^*(\tau,t_f,k_f)$ satisfy the optimal control equations for the problem with the boundary conditions:

$$k(t_0) = k_0,\ k(t_f) = k_f.$$

The name of *sensitivity analysis* (see Hadley & Kemp [1971]) is given to the problem of computing the first-order partial derivatives of J. We proceed directly:

$$\frac{\partial J}{\partial k_f}(t_f,k_f) = \int_{t_0}^{t_f} d\tau\{F_k \frac{\partial k^*}{\partial k_f} + F_c \frac{\partial c^*}{\partial k_f}\}$$

Now we may use the optimal control equations to note that:

$$F_c + \lambda f_c = 0 \text{ (maximum principle: H is maximized by c*)}$$

and

$$\dot{\lambda} = -\frac{\partial H}{\partial k} = -F_k - \lambda f_k.$$

Then

$$\frac{\partial J}{\partial k_f}(t_f,k_f) = -\int_{t_0}^{t_f} d\tau\{(\dot{\lambda} + \lambda f_k)\frac{\partial k^*}{\partial k_f} + \lambda f_c \frac{\partial c^*}{\partial k_f}\}$$

$$= -\int_{t_0}^{t_f} d\tau\{\dot{\lambda}\frac{\partial k^*}{\partial k_f} + \lambda \frac{df}{dk_f}\} \tag{17}$$

where by $\frac{df}{dk_f}$ is meant the total derivative of $f(k^*(\tau,t_f,k_f), c^*(\tau,t_f,k_f),\tau)$ with respect to k_f. The term $-\int_{t_0}^{t_f} d\tau\ \lambda\ \frac{\partial \dot{k}^*}{\partial k_f}$ can be integrated by parts to yield:

$$- [\lambda \frac{\partial k^*}{\partial k_f}]_{t=t_0}^{t=t_f} + \int_{t_0}^{t_f} d\tau\ \lambda\ \frac{\partial \dot{k}^*}{\partial k_f}$$

$$= - \lambda(t_f) + \int_{t_0}^{t_f} d\tau\ \lambda\ \frac{df}{dk_f} \tag{18}$$

The last step follows for these reasons: The value of $k^*(t_f,t_f,k_f)$ is k_f by definition, and the value of $k^*(t_0,t_f,k_f)$ is k_0. Hence $\frac{\partial k^*}{\partial k_f}(t=t_f) = 1$; $\frac{\partial k^*}{\partial k_f}(t=t_0) = 0$. Then the optimal path k^* satisfies $\dot{k}^*(\tau) = f(k^*,c^*,\tau)$, so that $\frac{\partial \dot{k}^*}{\partial k_f}$ is just the derivative with respect to k_f of $f(k^*(\tau,t_f,k_f), c^*(\tau,t_f,k_f),\tau)$, that is, $\frac{df}{dk_f}$ as defined above. Thus, finally:

$$\frac{\partial J}{\partial k_f}(t_f,k_f) = -\lambda(t_f) \tag{19}$$

This result is hardly surprising. Since $\lambda(t)$ was defined to be $\frac{\partial J^*}{\partial k_t}(k_t,t)$ for any time t between t_0 and t_f, it was interpreted as the marginal worth of stock at time t. What equation (19) says is that if the terminal constraint k_f is increased by one unit, then the value, J, of the optimal programme is decreased by the marginal worth of stock at the terminal time, t_f. It could hardly have been otherwise: if one relaxes the terminal constraint k_f, the benefit from the relaxation (at the margin) must be just the marginal benefit of stock at time t_f.

Now the economic interpretation of the other partial derivative of J, *viz.* $\frac{\partial J}{\partial t_f}$, is clear. It will turn out to be the marginal worth of *time* at the endpoint of the programme. The *transversality conditions,* which are defined to be the requirements which must be fulfilled by the optimal choices of k_f and t_f will thus be nothing other than the requirements that the marginal worth of stock at the end of the programme and the marginal worth of time at the end of the programme should both be zero: a very intuitively satisfying economic requirement.

What then is the marginal worth of time at the end of the programme? It is $\frac{\partial J}{\partial t_f}(t_f,k_f)$ which from the definition of J is just:

$$\frac{\partial J}{\partial t_f}(t_f,k_f) = F(k^*(t_f,t_f,k_f),\ c^*(t_f,t_f,k_f),t_f)$$

$$+ \int_{t_0}^{t_f} d\tau\{F_k\frac{\partial k^*}{\partial t_f} + F_c\ \frac{\partial c^*}{\partial t_f}\}\ .$$

By an exactly similar argument to that which led to equation (18), one obtains:

$$\frac{\partial J}{\partial t_f}(t_f,k_f) = F(t_f) - [\lambda\ \frac{\partial k^*}{\partial t_f}]_{t_0}^{t_f}$$

$$+ \int_{t_0}^{t_f} d\tau(\lambda\ \frac{df}{dt_f} - \lambda\ \frac{df}{dt_f}),$$

where $F(t_f)$ is interpreted obviously, as the value of F at time t_f along the optimal path. Clearly $\frac{\partial k^*}{\partial t_f}\ (t=t_0) = 0$. Since $k^*(t_f,t_f,k_f) = k_f$ by definition, it follows that

$$k_1^*(t_f,t_f,k_f) + k_2^*(t_f,t_f,k_f) = 0$$

where $k^*_{1,2}$ denotes the partial derivative of k* with respect to its 1st, 2nd argument. But $k^*_1 = \dot{k}^* = f(k^*,c^*,t)$, so that $\frac{\partial k^*}{\partial t_f}(t=t_f)$, which is just $k^*_2(t_f,t_f,k_f)$, equals $- f(k^*(t_f),c^*(t_f),t_f)$ (where some arguments of k* and c* have been omitted). Thus:

$$\frac{\partial J}{\partial t_f}(t_f,k_f) = F(t_f) + \lambda(t_f)f(t_f)$$

(in obvious notation)

$$= H(t_f) \tag{20}$$

In other words, the marginal worth of time at time t_f is just the Hamiltonian at time t_f. The two transversality conditions to be satisfied by an optimal choice of t_f and k_f are

$$\lambda(t_f) = 0; \qquad H(t_f) = 0 \tag{21}$$

two equations for two unknowns.

It can happen frequently that k_f and t_f are neither completely specified nor completely free to be chosen. It may be that some relation must be satisfied by them, say $S(t_f,k_f) = 0$. This modification is easily handled. The problem has become:

maximise $J(t_f,k_f)$ subject to $S(t_f,k_f) = 0$.

The Lagrangian is:

$$L = J(t_f,k_f) - p\, S(t_f,k_f)$$

and the first-order conditions for a maximum are therefore

$$\frac{\partial J}{\partial t_f} - p\,\frac{\partial S}{\partial t_f} = 0 \quad : \quad H(t_f) = p\,\frac{\partial S}{\partial t_f} \tag{22}$$

$$\frac{\partial J}{\partial k_f} - p\frac{\partial S}{\partial k_f} = 0 \; ; \quad -\lambda(t_f) = p\frac{\partial S}{\partial k_f} \; ;$$

$$S(t_f, k_f) = 0.$$

That's all! But it is perhaps worthwhile to clear up a common misconception of what is meant by the transversality conditions.

Figure 1 depicts an optimal path, $k^*(t)$, ending at the optimal point (t_f, k_f) of the curve whose equation is $S(t_f, k_f) = 0$.

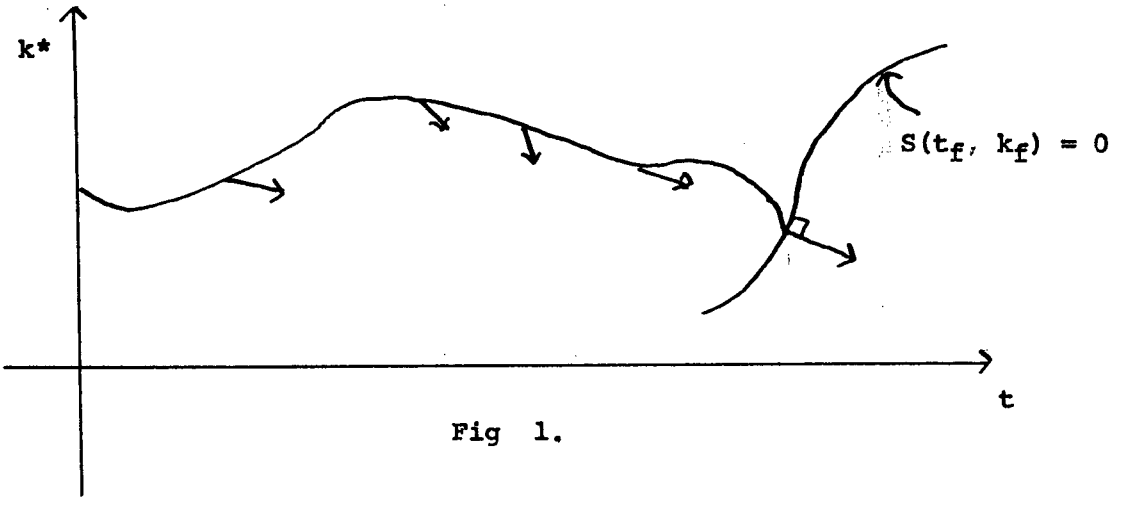


Fig 1.

As drawn, the path $k^*(t)$ does not intersect the curve S at right angles, and in general there is no reason for it to do so (in physics sometimes there is - hence the confusion). But at each point on the path, a little arrow is drawn, and this points in the direction of the vector $(H(t), -\lambda(t))$. The directions are as drawn if both H and λ are positive, as will often be the case in applications. This vector is indeed at right angles (transverse) to the curve S. Why so? This is just the content of eq (22), which tells us that the vectors $(H, -\lambda)$ and (S_t, S_k) are parallel at (t_f, k_f). (Partial derivative notation). But (S_t, S_k) is the gradient vector of the

function S (∇S as it is frequently written), and the gradient vector is the normal to the line of constant S, i.e., the curve $S(t_f, k_f) = 0$.

This kind of reasoning is perfectly familiar to engineers and physicists, but perhaps less so to economists. They would more likely reason as follows: the function $J(t_f, k_f)$ is just the maximum worth (derived utility perhaps) from a programme ending at (t_f, k_f). We are interested in maximising it subject to a constraint, *viz.* $S(t_f, k_f) = 0$. Thus usual procedure calls for us to draw indifference curves, that is, loci of points yielding the same utility J, and then to find where an indifference curve is tangent to the line of constraint. This is shown in Figure 2.

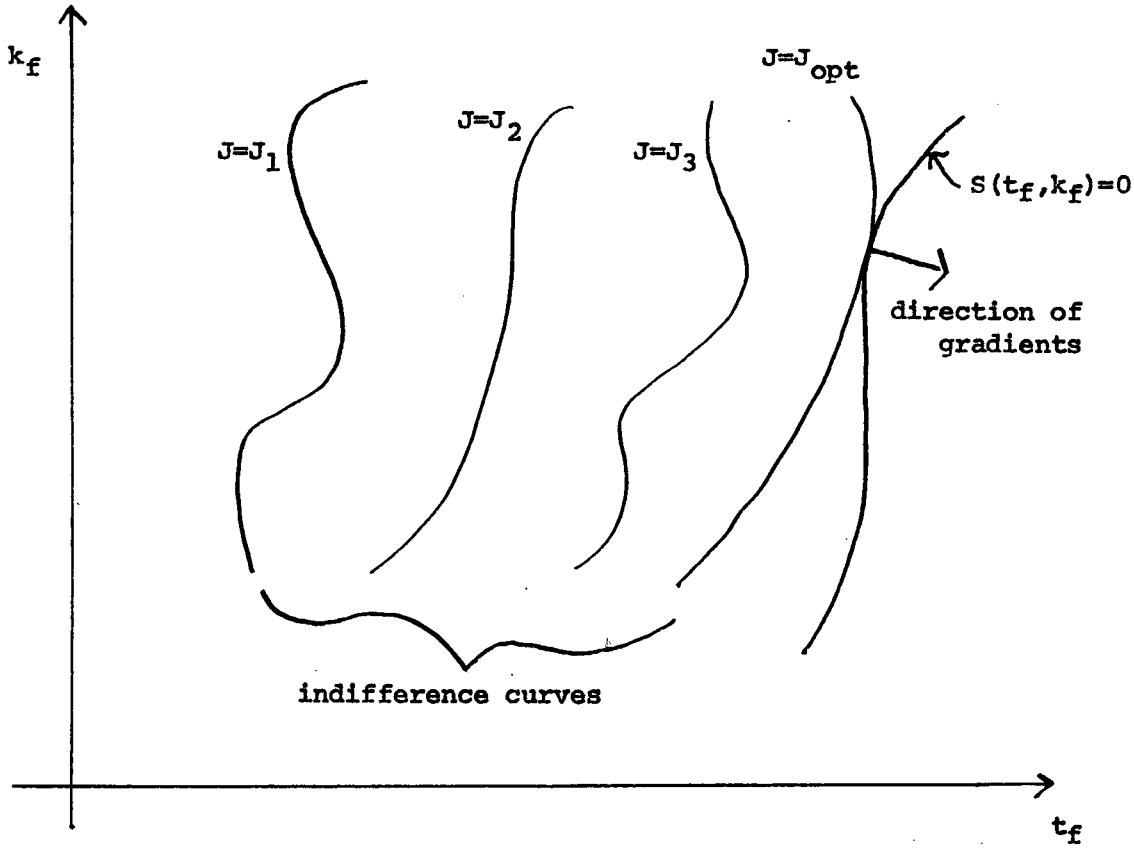


Fig 2.

At the tangency point, the gradients of $J(t_f,k_f)$ and $S(t_f,k_f)$ coincide, i.e., $\nabla J = \nabla S$, or more explicitly:

$$\frac{\partial J}{\partial t_f} = p\,\frac{\partial S}{\partial t_f} \quad ; \qquad \frac{\partial J}{\partial k_f} = p\,\frac{\partial S}{\partial k_f}; \quad S = 0$$

These are just the transversality conditions (22). Maybe if economists had created optimal control theory, they would be called the tangency conditions.

A few more remarks. The usual notions of quasiconcavity (convexity) can obviously be brought into play here so as to obtain *sufficient* conditions for a maximum. The requirement is easy to impose on S, but a bit trickier on J. Often the constraint S takes the form of fixing t_f or k_f while leaving the other variable open to choice. Then S will be something like $k_f - \bar{k}_f = 0$, say. The transversality conditions becomes simply (for this example)

$$H(t_f) = 0,\ k_f = \bar{k}_f.$$

(The other equation, $-\lambda(t_f) = p$ has no content, since p, too, is unknown). There is still enough to determine the problem fully: one equation (H=0) for one unknown (t_f).

4. The Marginal Worth of Time

The nature of time is a considerable mystery even to physical scientists. The shift from statics to dynamics in economic analysis involves difficulties whose source is precisely the nature of time. This essay certainly does not claim to solve all these difficulties. But I think it is fair to say that certain economic insights can be had from the considerations of optimal control theory, which are helpful precisely in the search for better understanding of intertemporal economic problems.

It was pointed out as unsurprising that the co-state variable λ should, as a Lagrange multiplier, be interpreted as a price and a marginal worth -- of stock, that is, the state variable. But sensitivity analysis revealed that the worth of time at the end of an optimal programme was the Hamiltonian, and that is not so **unsurprising. Again (p. 21)**, it was seen that the Hamiltonian was also the rate of acquisition of benefits -- the benefits accruing *during* one unit of time. This interpretation is not exactly the same as calling the Hamiltonian the marginal worth of time, but it is close. Moreover, it holds for *all* times between t_0 and t_f, not just at t_f. Yet again, the Hamilton-Jacobi-Bellman equation [eq. (16)] can be interpreted in this vein. Because of equation (14), equation (6) can be written as

$$J_2^*(k_t,t) = -H(k_t,c_t,\lambda_t,t) \tag{23}$$

in obvious notation. Now J* is not quite the same as J, being a function of different arguments. To clear this matter up, let us now include all the arguments of the maximised functional, and write

$$J^*(t,k_t,t_f,k_f) = \max_c \int_t^{t_f} d\tau F(k(\tau),c(\tau),\tau)$$

subject to

$$\dot{k} = f(k,c,t)$$

$$k(t) = k_t; \qquad k(t_f) = k_f$$

Then equation (23) can be written

$$\frac{\partial J^*}{\partial t} = - H(t)$$

and it is clear that this result, too, is saying something about the Hamiltonian and the marginal worth of time. But what exactly?

It says that if, at time t in the course of an optimal programme, one skips a unit of time, without anything else changing -- k stays the same, as well as t_f and k_f — and then proceeds optimally after this skip, then the *loss* of total benefit is H, evaluated at time t. In this precise sense, then, H is the marginal worth of time, at time t.

If an intertemporal maximisation is being carried out, and it may be a much more general one than that specified by equations (1), (2), (3), time is an input, or factor of benefit, more or less like any other. I must say "more or less", because it is only in some regards that it is like any other. It has a marginal worth, or price, certainly, and H, the Hamiltonian, comes close to measuring it. But in what circumstance does it measure what an economic agent would be willing to *pay* for a unit of time? Each instant of time is unique, with its own properties -- it is a *heterogeneous* input, and plainly at some times one would pay much for a few golden moments like the ones just experienced (excuse the language, but time *is* a mystery and therefore liable to provoke mystical talk). The difficulty, all too well known, is that moments of time, in their full individuality, cannot be either skipped or replicated, and so in no conceivable market could anyone receive, in exchange for whatever payment, an extra unit of time, at time t, valued to be sure at just H(t).

The physical world is not like this. The laws of motion are immutable, and what can happen at time t can happen at time t^1, for all t and t^1. (This principle, the homogeneity of time, was stated in a clear, and false, form by Newton, and once it was corrected and cast into a better

form by Einstein became a major ingredient of the Principle of Relativity). The physical world deals with inanimate objects, drab, dull and neither (*pace* Samuelson) profit maximising nor altruistic. Thus, for all these reasons the Hamiltonian, as defined by physicists for physical purposes, is quite exactly the marginal worth of time -- or as they would say, the infinitesimal generator of time translations, when these are taken as elements of the Poincaré group.

In economics, it is sometimes reasonable to maximise one's benefits as if time were homogeneous, and the distant future as vital as the present. Ramsay [1928], after all, told us it was *immoral* to discount. When this is so, the Hamiltonian is the marginal worth of time, as will be seen in a moment.

What an economic agent can perfectly well buy in an appropriate market is an extra unit of time in which to complete his programme. A student may bribe an invigilator for an extra few minutes to finish writing an exam; a big firm will bribe the government for more time to comply with anti-pollution laws; options can quite legally be written into contracts to allow a contracting party more time, at a price, to carry out his obligations; most familiar of all, payment of interest will buy time to repay a debt. Where does all this sound economic sense appear in optimal control theory? The answer is easy: as $\frac{\partial J^*}{\partial t_f}(t,k_t,t_f,k_f)$. This is the worth of an extra unit of time *tacked on* at the end of the programme. That unit is what it is, and may be quite different from any preceding or subsequent unit, but it is a definite, unmystical unit, and its worth is easily determined, as $\partial J^*/\partial t_f$.

At this point, economic science gives us a very large bonus. We all have the feeling that when a number of inputs contribute to output or

utility or whatever -- benefit in general -- then the optimal result is achieved when the quantities of the inputs are chosen so that the extra worth deriving from a unit of further expenditure on any one of the inputs is the same as that from any other. Time is only one input, but a heterogeneous one, and so each instant is like a separate input. The result which corresponds to the above familiar one is this: it does not matter at what time t an economic agent purchases an extra unit of time to be tacked on at the end; the extra, or marginal, worth it provides will always be the same, along an optimal path.

This result is not exactly the same as the usual one about inputs, but it is just as useful, and it is proved in the same way. For, because J* is defined as an *integral*, we can write,

$$J^*(t_0,k_0,t_f,k_f) = J^*(t_0,k_0,t,k_t^*) + J^*(t,k_t^*,t_f,k_f)$$

where k_t^* is the value, $k^*(t)$ of k along the optimal path from t_0 to t_f, at time t. Now consider a slightly longer programme, lasting till $t_f + dt$. *If* one were constrained to pass through the point (t,k_t^*) on the new trajectory, the (constrained) maximum worth to be had would be

$$J^*(t_0,k_0,t,k_t^*) + J^*(t,k_t^*,t_f + dt,k_f)$$

$$= J^*(t_0,k_0,t_f,k_f) + \frac{\partial J^*}{\partial t_f}(t,k_t^*,t_f,k_f)dt + o(dt)$$

But this must be less than the unconstrained maximum, which is

$$J^*(t_0,k_0,t_f + dt,k_f) = J^*(t_o,k_0,t_f,k_f) + \frac{\partial J^*}{\partial t_f}(t_0,k_0,t_f,k_f)dt + o(dt)$$

Therefore $\frac{\partial J^*}{\partial t_f}(t,k_t^*,t_f,k_f) \leq \frac{\partial J^*}{\partial t_f}(t_0,k_0,t_f,k_f)$. If one considers a *shorter*

programme, of length t_f-dt, it is plain that:

$$J^*(t_0,k_0,t_f,k_f) - \frac{\partial J^*}{\partial t_f}(t,k_t^*,t_f,k_f)dt \leq J^*(t_o,k_0,t_f,k_f) - \frac{\partial J^*}{\partial t_f}(t_0,k_0,t_f.k_f)dt$$

whence

$$\frac{\partial J^*}{\partial t_f}(t,k_t^*,t_f,k_f) \geq \frac{\partial J^*}{\partial t_f}(t_0,k_0,t_f,k_f)$$

But the two results show that

$$\frac{\partial J^*}{\partial t_f}(t,k_t^*,t_f,k_f) = \frac{\partial J^*}{\partial t_f}(t_0,k_0,t_f,k_f) \tag{24}$$

for all t. And this is the desired result. The extra worth from purchasing an extra unit of time, at time t [left-hand side of equation (24)] is the same for all times t between t_0 and t_f.

The argument is just as easy in words. The gain in worth from being given an extra unit of time at t cannot be more than that from being given it at t_0, or else the path from t_0 to t_f through k_t^* could not be optimal. Delayed information (i.e., greater length of time available) cannot be more valuable than the same information provided earlier. On the other hand, the loss from being deprived of a unit of time at time t must be at least as great as that of being deprived of it at t_0 for exactly the same reason: delay of bad news cannot make things any better. But at the margin, the gain from a unit of time is the same as the loss from being deprived of it, and so both are equal to the gain from the extra unit purchased at t_0, and therefore constant along the optimal path. This result, like the familiar one about inputs, characterises *optimal* paths, and is proved by exploiting optimality.

I said above that this result was a large bonus. I shall now give two examples which will, I trust, vindicate that statement. The first

is the example of *autonomous* systems, that is, those drab, dull ones like physical systems or Ramsay's moral ones, where one instant of time is exactly like any other. What this means is that the functions F and f of equations (1) and (2) cannot depend explicitly on t. Consequently, neither does H, which is nothing but $F + \lambda f$. In fact, since

$$J^*(t,k,t_f,k_f)$$

$$= \max\int_t^{t_f} F(k(\tau),c(\tau))d\tau$$

[subject to $\dot{k} = f(k,c)$ and $k(t) = k, k(t_f) = k_f$]

$$= \max\int_{t+t'}^{t_f+t'} F(k(\tau),c(\tau))d\tau \tag{25}$$

[subject to $\dot{k} = f(k,c)$ and $k(t+t') = k, k(t_f+t') = k_f$],

we have $J^*(t+t',k,t_f+t',k_f) = J^*(t,k,t_f,k_f)$ or else, more simply, $J^*(t,k,t_f,k_f)$ is a function of k_0, k_f and t_f-t alone. (That is, time is *homogeneous*, and only time *differences* matter: equation (25) says that if both t and t_f are translated an equal amount, then J^* is not changed). But then, the result we have obtained, namely that $\frac{\partial J^*}{\partial t_f}$ is constant along the optimal path, means also that $-\frac{\partial J^*}{\partial t}$, i.e., H(t), is constant along the optimal path, since, if $J^*(t,t_f) = P(t_f-t)$, say, suppressing k - dependences, it is immediate that $\frac{\partial J^*}{\partial t_f} = -\frac{\partial J^*}{\partial t}$. We have shown in this case that H(t) is indeed the marginal worth of time. Now, on p. (19), it was pointed out that once the maximum principle, equation (14), was used, equations (11) and (12) were two first-order differential equations for k and λ. Knowledge that H = constant is the same as knowledge of a first integral of these equations. In other words, if one substitutes $c^* = c^*(k,\lambda)$ from equation (14) into $H(k,c^*,\lambda)$ = constant, one

obtains the equation of the trajectories generated in the *phase* plane [k,λ)-space] by the equations (11) and (12). These are usually called the *optimal trajectories*. If one decides to look at one of them, that is if one fixes the value of H, then one may solve $H(k,c^*(k,\lambda),\lambda) = \bar{H}$ for either k or λ, substitute the solution into equation (11) or equation (12) and get a single differential equation with just one unknown function of time. For example, one may obtain $\dot{k} = h(k)$, say. This equation can always be solved by integrating $\frac{1}{h(k)}$, as follows:

$$t - t_0 = \int_{k_0}^{k} \frac{dk'}{h(k')}.$$

(The solution has been reduced to *quadrature*, as the old books on applied mathematics say). This very pleasing result is quite general if there is only one state variable, k. It is not so useful in the case of several state variables (a *vector* k of state variables) each with a corresponding co-state variable, but it is not entirely worthless either.

The second example is the case where the heterogeneity of time enters only because of Ramsay immorality, that is, discounting of future benefits. Here, $\partial J^*/\partial t_f$ turns out to be an interesting quantity although it is no longer H. The Hamiltonian can be written as:

$$H(k,c,\lambda,t) = e^{-\delta t}\hat{F}(k,c) + \lambda f(k,c),$$

so that the optimal control equations are as follows:

$$\frac{\partial H}{\partial c} = e^{-\delta t}\hat{F}_c(k,c) + \lambda f_c(k,c) = 0$$

$$\dot{\lambda} = -\frac{\partial H}{\partial k} = - e^{-\delta t}\hat{F}_k - \lambda f_k$$

$$\dot{k} = \frac{\partial H}{\partial \lambda} = f(k,c). \qquad (26).$$

If now one puts $\lambda = pe^{-\delta t}$, the resulting system of equations in k and p (and c) is autonomous:

$$\hat{F}_c(k,c) + p\, f_c(k,c) = 0$$

$$\dot{p} - \delta p = - F_k - pf_k$$

$$\dot{k} = f(k,c) \tag{27}$$

As in the fully autonomous case, if a first integral of this system can be found, the problem is essentially solved. But H is not a first integral.

Before deriving what the first integral is, it is worthwhile to spend just a few moments more on H. It is not constant, and its total time derivative can be calculated:

$$\frac{dH}{dt}(k,c,\lambda,t) = \frac{\partial H}{\partial t} + \frac{\partial H}{\partial k}\dot{k} + \frac{\partial H}{\partial c}\dot{c} + \frac{\partial H}{\partial \lambda}\dot{\lambda}$$

$$= \frac{\partial H}{\partial t} + \dot{\lambda}\dot{k} - \dot{k}\dot{\lambda}, \text{ by equations (26)}$$

$$= \partial H/\partial t.$$

All the time dependence of H is the explicit part, everything else cancels out along an optimal path. This calculation provides an alternative proof of the constancy of H in autonomous systems.

The real constant is $\partial J^*/\partial t_f$. We may observe that

$$J^*(t+dt,k_t,t_f,k_f)$$

$$= \int_{t+dt}^{t_f} e^{-\delta\tau}\hat{F}(k^*(\tau),c^*(\tau))d\tau \tag{28}$$

where $k^*(\tau)$ and $c^*(\tau)$ define the optimal path and where in particular

$$k^*(t+dt) = k_t \text{ and } k^*(t_f) = k_f.$$

The integral of equation (28) can also be written, after changing the dummy variable of integration τ to $\tau' + dt$, as

$$e^{-\delta(dt)}\int_t^{t_f-dt} e^{-\delta\tau}\hat{F}(k^*(\tau'+dt),c^*(\tau'+dt))d\tau'$$

Similarly,

$$J^*(t,k_t,t_f-dt,k_f) = \int_t^{t_f-dt} e^{-\delta\tau}\hat{F}(\hat{k}(\tau),\hat{c}(\tau))d\tau$$

where $\hat{k}$ and $\hat{c}$ define the optimal path with boundary conditions

$$\hat{k}(t) = k_t, \quad \hat{k}(t_f-dt) = k_f$$

But $\hat{k}(\tau)$ must equal $k^*(\tau+dt)$ and similarly $\hat{c}(\tau)$ must equal $c^*(\tau+dt)$, since the hatted and starred pairs of functions are defined by the same *autonomous* system, equation (27), with the same boundary conditions once the translation by dt has been attended to for k^* and c^*.

Hence

$$J^*(t+dt,k_t,t_f,k_f) = e^{-\delta(dt)}J^*(t,k_t,t_f-dt,k_f).$$

This yields

$$J^*(t,k_t,t_f,k_f) + dt\cdot\frac{\partial J^*}{\partial t} = e^{-\delta(dt)}[J^*(t,k_t,t_f,k_f) - dt\cdot\frac{\partial J^*}{\partial t_f}] + o(dt)$$

$$= J^*(t,k_t,t_f,k_f) - dt[\frac{\partial J^*}{\partial t_f} + \delta J^*] + o(dt)$$

and thus

$$\frac{\partial J^*}{\partial t_f} = -\frac{\partial J^*}{\partial t} - \delta J^*$$

$$= H - \delta J^*. \qquad \text{[Equation (23)]}$$

So that the constant along an optimal trajectory, and also the marginal worth of time, is $H - \delta J^*$. This result is useful in a quite different way from the autonomous system result. There H = constant gave the equation of the optimal trajectories, but here, since J^* is what is to be *found* by solving the optimal control problem rather than being a *known* function, $H - \delta J^*$ = constant is of no help in finding the equation of the trajectories. But these after all are given by the autonomous system (27), which may yield a first integral by direct methods. Then, if that is so, the trajectories are known, and consequently H can be calculated along them directly. At the *end* of any trajectory, $J^* = 0$ by definition, and so the constant appropriate to that trajectory is known. In particular, if the end of the trajectory is characterised by the transversality condition $H = 0$, the constant is zero. This means that J^* itself, the worth of the optimal programme, can be obtained at any point on a trajectory, without fully solving the problem and carrying out the integration of equation (1). This information can be very valuable. The case of the $H = 0$ transversality condition is especially interesting. For then we get the very agreeable economic result that, for an optimal programme for which time is not a constraint -- and so is chosen optimally -- the remaining value of the programme, at any point in its course, is obtained simply by solving

$$H - \delta J^* = 0, \text{ i.e., } \quad J^* = \frac{1}{\delta} H;$$

in other words simply by taking the capitalised value of the Hamiltonian at the given rate of discount, δ. The remaining worth, J^*, is equivalent to a perpetual benefit stream of size H, at interest rate δ.

These last results will be of great use in the other two essays of this thesis.

5. The Phase Plane

In this final section, let us restrict our attention to problems which are either autonomous or time-dependent only through an exponential discounting factor $e^{-\delta t}$. Equation (27) gives the autonomous optimal control equations for k and the "current-value" shadow price p:

$$\hat{F}_c(k,c) + pf_c(k,c) = 0$$

$$\dot{p} = -\hat{F}_k - pf_k + \delta p$$

$$\dot{k} = f(k,c) \tag{27}$$

Let the solution of the first of these equations (the maximum principle) be written as

$$c = c(k, p) \tag{28}$$

This may now be substituted into the remaining two equations to yield a closed system. The *phase plane* of the problem is constructed by taking k as abscissa, p as ordinate, plotting the two lines $\dot{p}=0$, $\dot{k}=0$, and drawing the trajectories of the solutions $k(t)$, $p(t)$ of equations (27) for varying initial conditions. Frequently the result will resemble Figure 3.

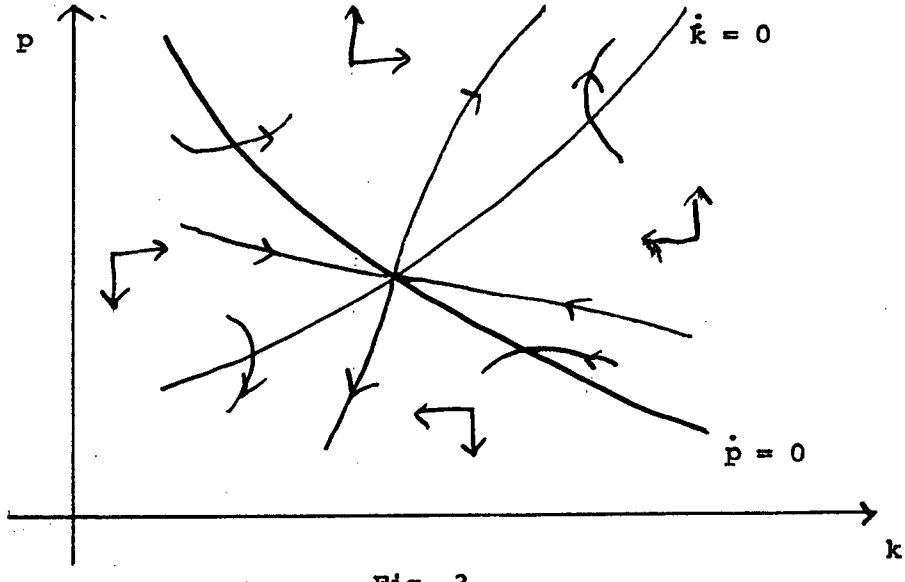


Fig 3.

The lines $\dot{k} = 0$ and $\dot{p} = 0$ can be best interpreted by imagining a *static* world. First of all, if an economic agent could perceive only a spot price, p, of stock, and further presumed that this price would last for ever, he would wish to purchase an amount, k, of the stock such that its marginal worth equalled p - that is, such that the marginal unit yielded zero net profit. How much benefit does a stock k, costing pk, yield? In a short space of time, Δt say, if our agent chooses a value c of the control variable, the direct benefit is $\hat{F}(k, c)\Delta t$. The stock has changed by an amount $\dot{k}\Delta t$, that is $f(k, c)\Delta t$, which, since the price stays fixed at p, is worth $pf(k, c)\Delta t$. The original stock is still worth its cost pk, but interest has been forgone on this sum to the value of $\delta pk\Delta t$. The net benefit is thus $(\hat{F}(k, c) + pf(k, c) - \delta pk)\Delta t$.

The first-order conditions for maximising this are just

$$\hat{F}_c + pf_c = 0 \; ; \quad \hat{F}_k + pf_k - \delta p = 0,$$

and it is immediate from equations (27) that this yields the $\dot{p}=0$ line in the phase plane. Thus this line can be interpreted as the static (derived) *demand curve* for stock.

Next, one can imagine a holder of stock, rather than a purchaser. If his holding is k, what would it cost for him to produce an extra unit of stock? The cost will be the difference in benefit received over a short time Δt in the following two sets of circumstances:

(a) the stock k is maintained unchanged during Δt,

(b) the stock k is increased by one unit during Δt.

For (a), the control variable c must be chosen so that $\dot{k} = 0$, that is $f(k,c) = 0$. Let the solution to this equation be $c = c(k)$. Benefit received is thus $F(k,c(k))\Delta t$. For (b) the control variable must be chosen as $c(k) + \Delta c$, say, where $\dot{k}\Delta t = 1$, that is $f(k,c(k) + \Delta c)\Delta t = 1$. Since $f(k,c(k)) = 0$, this means that

$$\Delta c \Delta t = 1/f_c(k,c(k)) \tag{29}$$

The benefit received is then $F(k,c(k) + \Delta c)\Delta t = F(k,c(k))\Delta t + F_c(k,c(k))\Delta c\Delta t$. The benefit forgone in producing one unit of stock is, therefore, from equation (29):

$$-F_c(k,c(k))\Delta c\Delta t = -\frac{F_c(k,c(k))}{f_c(k,c(k))}$$

It is reasonable to call this quantity the *supply price* of stock, p_s, say. Then the variables p_s, k and c(k) satisfy the pair of equations

$$F_c(k,c) + \dot{p}\, f_c(k,c) = 0$$

$$f(k,c) = 0$$

, and from eqs (27). this gives precisely the $\dot{k} = 0$ line in the phase plane. This line can then be

interpreted as the static *supply curve* for stock.

These arguments show that, when the phase plane looks like Fig. 3, it can indeed be understood as an ordinary demand-and-supply diagram, with adjustment mechanisms provided. These mechanisms are the key to comparative dynamics, as will become clear in the examples of the remaining two essays of the thesis.

To conclude this essay, some comments are in order on the limitations of the analysis presented. Throughout, only one state variable, k, has been considered. Problems can of course easily arise in which two or more are needed -- natural resource models are the most obvious example; see, for example, the bibliography in Clark [1974]. Such problems, where naturally a two-dimensional phase plane is no longer sufficient, are harder to treat than the problem of this essay in the same measure as multi-product general-equilibrium models are harder than one-product partial equilibrium ones. On the other hand, where there is only one state variable, more than one control variable can be handled as easily as can one -- the model of the following essay is an example of such a case. The development and analysis of the optimal control algorithm, as well as of the transversality conditions, is applicable to a problem with a vector of state variables with virtually no modifications. It is the subsequent analysis, both mathematical and economic, that is difficult.

The technical limitations of the presentation of this chapter are manifest. It is enough to work through the book of Hadley and Kemp [1971] to become aware of the great variety of subtle mathematical points which can arise in optimal control questions. But Hadley andKemp state their aim as being to write a mathematical textbook with examples drawn from economics: mine has been the obverse -- to simplify the mathematics so as to clarify the economics.

CHAPTER II

A STOCK ADJUSTMENT MODEL AND THE PROBLEM OF OPTIMAL EXIT

This essay builds on the work of the preceding one. A specific model is treated which allows the economic interpretation of optimal control theory to be used for working out comparative dynamics. The results are not particularly difficult. This fact may seem surprising to anyone who has studied the work of Oniki (1973) and it may be worthwhile to ask why. Oniki has successfully (I believe) followed Samuelson's instructions for comparative statics in the case of comparative dynamics. That is, he has computed expressions for the (infinitesimal) changes induced at each instant in all the endogenous variables in an optimal control problem by changes in any of the exogenous variables. It is not surprising that these expressions are complicated in their general forms and, besides, usually impossible to sign. But often the interesting economic consequences of a change in exogenous variables are restricted to a small number of those calculated by Oniki. Again, as will become clear in this essay and the next, it can be the effects of a *finite* change in an economic environment which are of real interest rather than a tendency, expressed by some derivatives, produced by infinitesimal change. These remarks are in no way meant as a slight on Oniki's extremely valuable work (for he has a real compendium of results in very general form), but instead are intended to point out the substantial difference between his approach and mine. There are questions - of detail rather than of essence - that can quite legitimately be called questions in comparative dynamics

which I shall not consider at all in what follows, because I think it is more fruitful at the moment to concentrate on matters which are distinctively related to dynamics and time. Perhaps fortunately, perhaps because of my choice of models, these matters lend themselves to a reasonably simple treatment. The detailed questions left unanswered can, after all, be approached by Oniki's methods, which are chiefly an extension of those of comparative statics. The main point that I want to emphasize is that I shall be concerned with *non-infinitesimal* changes in exogenous variables.

In Section 1, the model of stock adjustment to be considered is specified. It can be thought of as providing an elementary paradigm for comparative dynamics, as I wish to look at it in this thesis. The optimal control equations are written down, and the phase plane drawn. Then in Section 2, a (finite) change is made in the cost structure for production of the stock. The dynamic adjustments made as a result are analysed, with special emphasis being given to the possibility of exit from business. The methods of Chapter 1 are used extensively in this section.

1. Specification of the Model

The question considered in this section is the following: if a firm keeps an inventory of finished goods out of which to meet demand for its product, what will be the effect, short-run and long-run, on the size of this inventory of an increase in production costs? We may observe at once that if only long-run effects are to be considered, the traditional comparative statics methods for long-run equilibria are sufficient to

answer the question. The model which is set up allows a definite qualitative answer to that static question and then the path the firm follows in going from its original situation to its final one is examined.

First, although optimal control theory allows treatment of situations where the firm is *not* in equilibrium before its costs rise, there are too many kinds of disequilibrium for an investigation of all of them to be useful. Consequently we shall assume that the firm starts in a position of long-run equilibrum. (Comparative *statics* certainly requires this).

The model: the firm's objective is taken to be the maximisation of the discounted sum of its profits into the indefinite future. Let the flow rate of sales be s, and let the flow of revenue from this sales volume be R(s). Let the flow rate of production be y, and the flow of cost from this be C(y). The functions R and C will be assumed to have the usual convenient properties of differentiability and diminishing returns. Let the inventory stock be denoted by k, and let costs of holding this stock be directly proportional to k, Θk, say. Then the discounted sum of profits is (discount rate = r) :

$$J = \int_0^\infty e^{-rt}[R(s) - \Theta k - C(y)]dt.$$

Output goes into inventory until sold, so that

$$\dot{k} = y - s.$$

It is necessary that $k \geq 0$, $y \geq 0$, $s \geq 0$. In order that there should be a purpose for holding inventory, the assumption is made that the sales flow cannot exceed some fraction of inventory:

$$s \leq qk. \tag{1}$$

This last constraint can be thought of as a crude model of the technology of distribution. If it takes a certain time for goods to be shipped from a factory to points of sale, then only a limited quantity determined by the size of inventory, can be used to restock shelves when their previous contents have been sold. The precise form of inequality (1) is in any case not critical to the results to follow. All that matters is that there should be some constraint on sales volume related to inventory size. Similarly it will become apparent in the analysis below that the direct cost, Θk, of holding inventory need not have that particular form, and in fact may be zero without affecting qualitative results, so long as the discount rate, r, is positive, since a carrying cost (forgone interest) is thereby introduced.

The Hamiltonian can now be formed. Following the rules of Chapter I, we obtain:

$$\begin{aligned} &H(k,p,s,y,t) \\ &= e^{-rt}[R(s) - \Theta k - C(y) + p(y-s)] \end{aligned}$$

where k is the state variable, p the (current value) co-state variable, s and y control variables. The maximum principle requires that we maximise H with respect to s and y for any admissible k and p over the feasible set of s and y:

$$0 \leq s \leq qk$$

$$y \geq 0.$$

The optimal values s* and y* are as follows:

$$s^* = \begin{cases} 0 & \text{if } p \geq R'(0) \\ s(p) & \text{if } R'(0) \geq p \geq R'(qk) \\ qk & \text{if } p \leq R'(qk) \end{cases}$$

$$y^* = \begin{cases} 0 & \text{if } p \leq C'(0) \\ y(p) & \text{if } p \geq C'(0) \end{cases}$$

where the functions s and y are the inverse functions of R' and C' and therefore satisfy the following identities:

$$R'(s(p)) = p$$

$$C'(y(p)) = p$$

The derivatives, R' and C', are presumed to be monotonic functions: R', the marginal revenue, decreasing, and C', the marginal cost, increasing. The maximised Hamiltonian is now:

$$M(k,p,t) = e^{-rt}(R(s^*) - \Theta k - C(y^*) + p(y^* - s^*)), \qquad (2)$$

and from this one may obtain the shadow-price equation:

$$\dot{p} = rp - e^{rt}M_k \qquad \text{[eq(I-27)]}$$

Five different regions of the phase plane can be distinguished and the forms of the shadow-price equation as well as of the stock equation $\dot{k} = y-s$ are listed below in the table, while the regions of the phase plane are depicted in Fig. 4.

TABLE 1

STOCK AND SHADOW-PRICE EQUATIONS

	p	y^*	s^*	$\dot{k} = y - s$	$p = rp - M_k e^{rt}$
I.	$p \geq R'(0)$	$y(p)$	0	$y(p)$	$rp + \theta$
II.	$p \leq R'(0)$ $p \geq R'(qk)$ $p \geq C'(0)$	$y(p)$	$s(p)$	$y(p) - s(p)$	$rp + \theta$
III.	$p \leq C'(0)$ $p \geq R'(qk)$	0	$s(p)$	$-s(p)$	$rp + \theta$
IV.	$p \leq C'(0)$ $p \leq R'(qk)$	0	qk	$-qk$	$rp + \theta + q(p-R'(qk))$
V.	$p \geq C'(0)$ $p \leq R'(qk)$	$y(p)$	qk	$y(p) - qk$	$rp + \theta + q(p-R'(qk))$

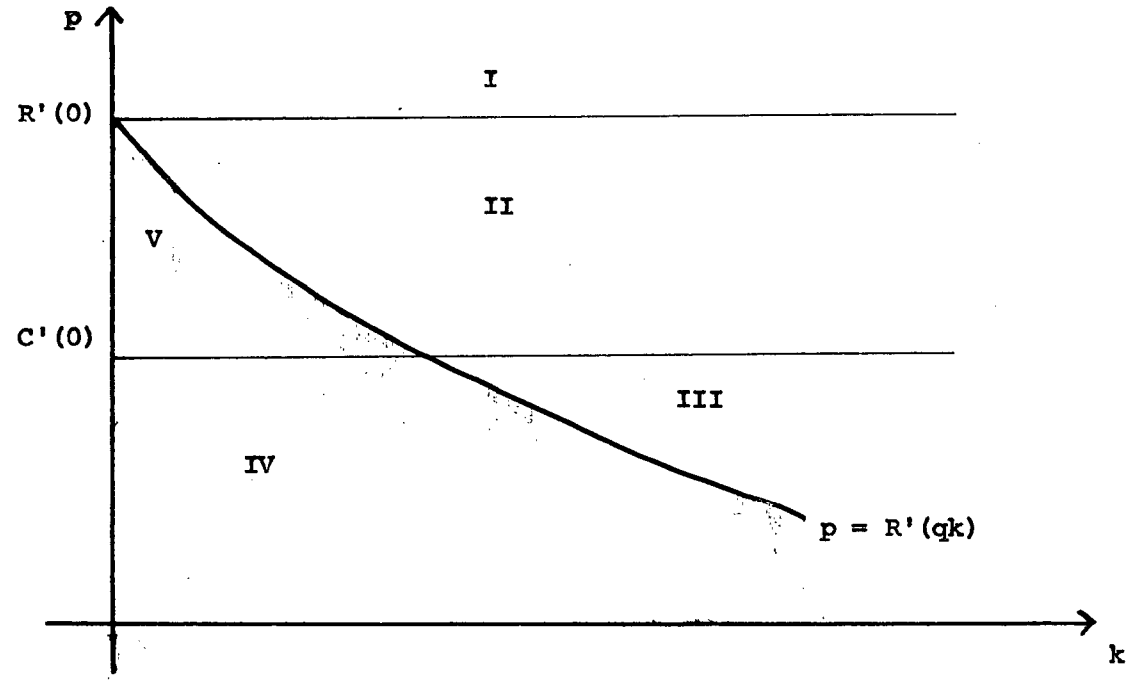


Fig 4.

(Unless $R'(0) > C'(0)$, production will never be profitable).

With this information, it is now possible to draw the $\dot{p} = 0$ (static demand) line and the $\dot{k} = 0$ (static supply) line. The result is in Fig 5, along with the senses of the optimal paths in the several parts of the phase plane.

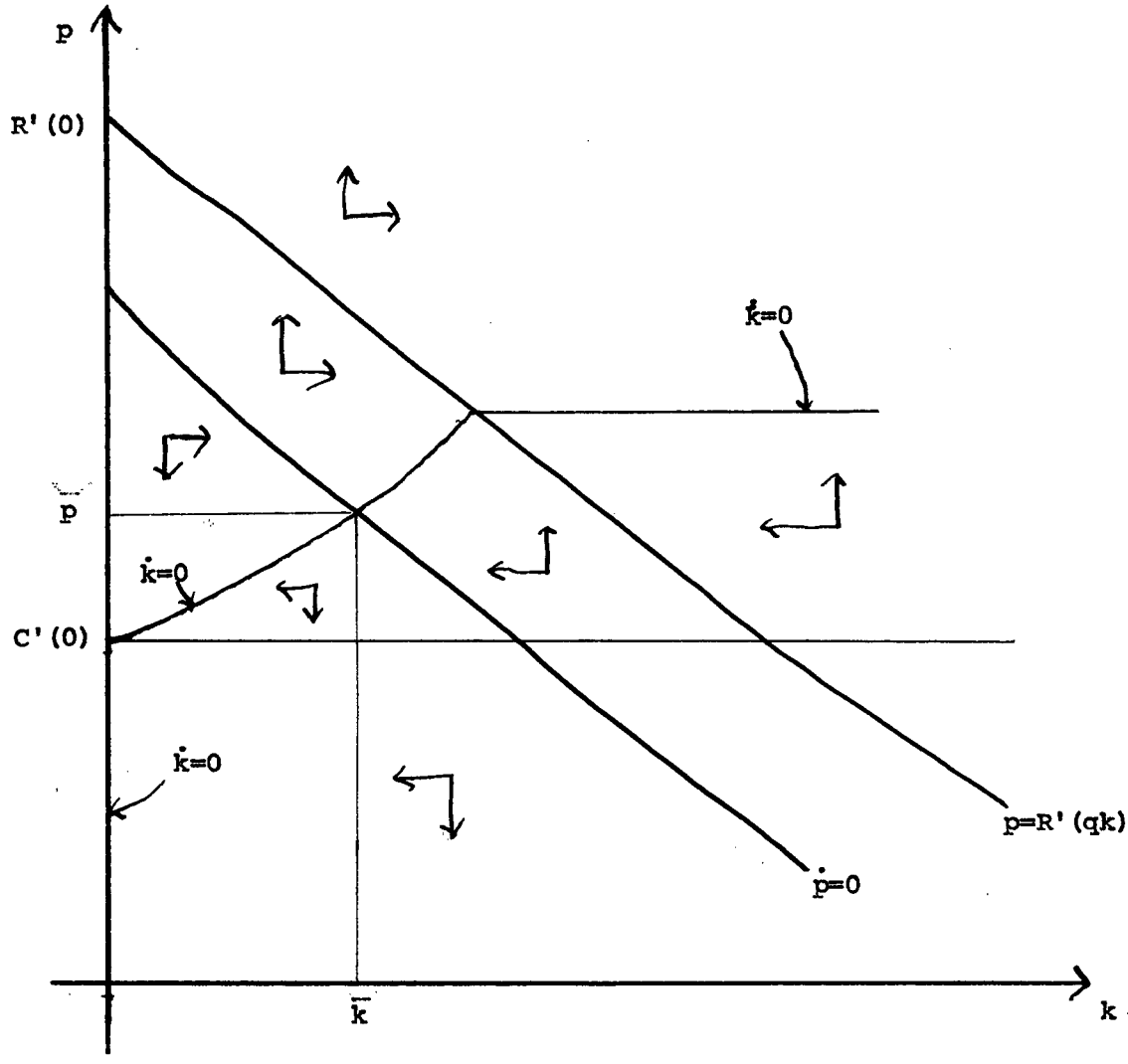


Fig 5.

The equation of the $\dot{p} = 0$ line is

$$p = \frac{qR'(qk)}{r+q} - \frac{\theta}{r+q}$$

which plainly must lie beneath the line $p = R'(qk)$, as drawn. The $\dot{k} = 0$ line is just the p-axis in region IV: in region V, it has equation $p = C'(qk)$; in Region II, it is the line $p = \tilde{p}$, where $\tilde{p}$ satisfies $y(\tilde{p}) = s(\tilde{p})$. It is evident that $\tilde{p}$ is the value of p where the curves $p = R'(qk)$ and $p = C'(qk)$ intersect.

It is possible to verify directly for this model that the $\dot{p} = 0$ and $\dot{k} = 0$ lines are respectively the static demand and supply curves. To do so it is convenient to imagine the firm divided into three departments, production, inventory and sales. Then the demand curve gives the prices (in static situations) that the sales manager is willing to pay the inventory manager for the latter's maintaining an inventory of a given size. The supply curve gives the prices that the production manager requires to be paid to supply goods sufficient to maintain inventories of given sizes. The analysis of Chapter 1, Section 5, makes it clear that the above statements are true but it may be illuminating to check them explicitly.

First, the supply curve (attention will be restricted to Region V: nothing of extra interest appears in the other regions.) In a static situation inventory is constant, and so in Region V, $s = y = qk$, all constants. We wish to know the marginal cost of a unit of stock in an inventory of size k. Such an inventory calls for a production rate, y, equal to qk. If one extra unit of stock is to be produced in one unit of time, then this calls for a production rate of y+1 for this unit of time and a marginal cost of

$C'(y) = C'(qk)$. It should be noted that this is the cost of the last unit of *stock* taken from the *flow* of production.

Next, the demand curve. This gives the net revenue achieved by the inventory manager from the sale of the last unit of stock to the sales manager. Let this revenue be denoted by p. Since sales volume is $s = qk$, the revenue obtained by the sales manager from one more unit sold is $R'(qk)$, and this then is what will be paid to the inventory manager for it. Against this, there are some charges to be borne by the inventory department. To provide one more unit (over unit time, say, although this condition does not affect the result) the inventory level must be raised by $\frac{1}{q}$ units because of the turnover constraint $s \leq qk$. The cost of this increase for one unit of time is $rp/q + \theta/q$: interest cost plus holding cost. Thus net revenue $= p = R'(qk) - rp/q - \theta/q$, whence $p = qR'(qk)/r+q - \theta/r+q$. This agrees with the equation of the $\dot{p} = 0$ line.

The point $(\bar{k},\bar{p})$ in Fig. 5 is the saddlepoint of the optimal paths in the phase plane. It should now be clear that it is indeed just the long-run equilibrium maintained by the inventory manager in our fictitious decentralised scheme of the firm. It is then the point that we assume the firm occupies when an exogenous cost increase takes place. The neighbourhood of this point is shown in Fig. 6, along with the stable and unstable arms leading to and from it, and the sense of the adjustment paths.

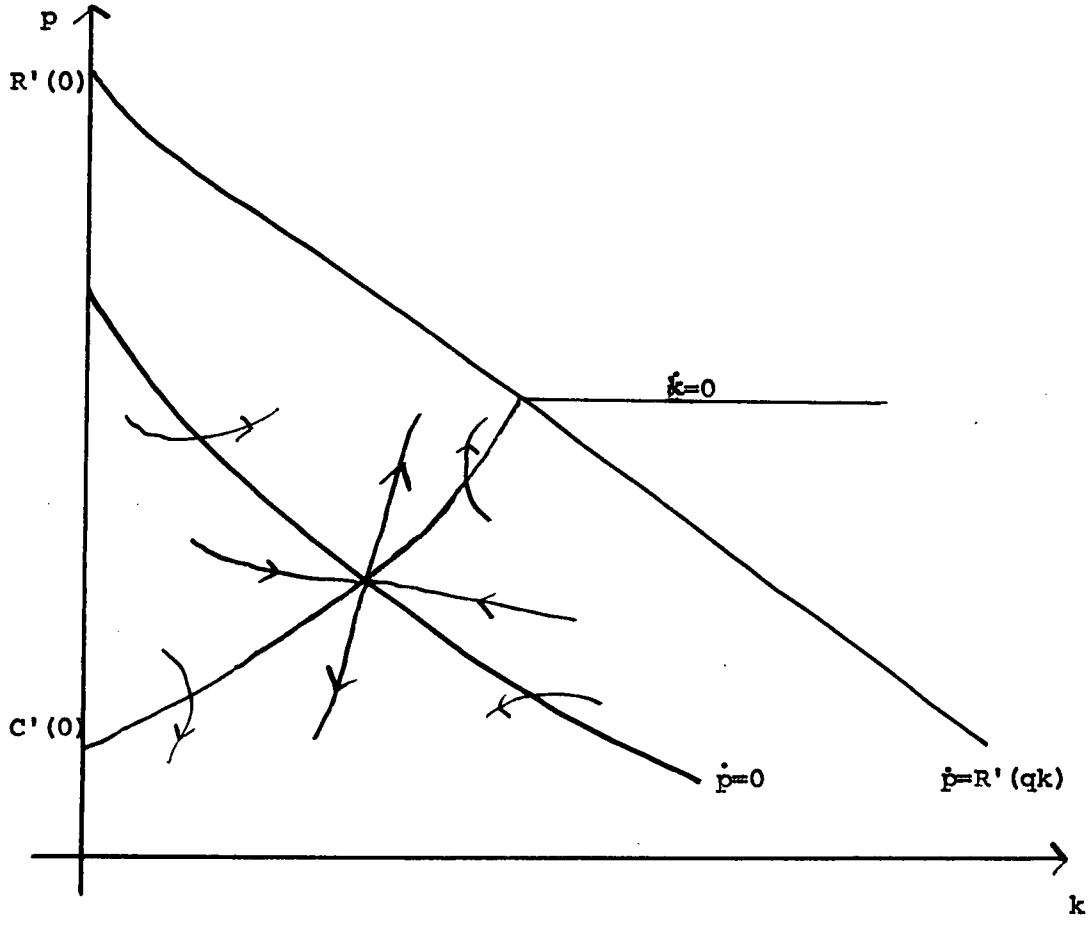


Fig 6.

2. Comparative Dynamics

Now that the optimal control problem has been worked out, we shall see that some comparative dynamics questions are not very difficult. There is assumed to take place a change in the cost function $C(y)$. It is not the function C itself but its derivative C', the marginal cost function, which defines the $\dot{k} = 0$ line in the phase plane (see Fig 6) and an increase in $C(y)$ for all y does not necessarily mean an increase (shift upwards) in $C'(y)$ -- an increase in fixed costs alone, for example, leaves $C'(y)$ unchanged. But $C'(y)$ will increase if the cost change is, for example, a specific tax on output produced or inputs used in production. Besides this is the usual state of affairs meant when one speaks of a supply curve being shifted up because of increased costs. Let us begin with this case. The long-run effect is no more difficult than the most elementary demand-and-supply analysis. The new long-run equilibrium (saddlepoint) lies on the $\dot{p} = 0$ line further up than the old $(\bar{k},\bar{p})$.

The next question is: Does the firm wish to adjust towards the new saddlepoint or go out of business? Let us for the moment assume that positive profits were being made at $(\bar{k},\bar{p})$ (i.e., our firm was intramarginal in its industry), so that if the cost change is small enough, it is still worthwhile to stay in business. Now the stock $\bar{k}$ is not instantaneously adjustable, but its shadow price may well change discontinuously with the cost increase (since we are considering a *finite* change in costs). Since the firm's optimum policy is to move to the new saddlepoint, we can tell in fact that it must move along a stable arm to get there. Thus immediately after the change, it goes to the point $(\bar{k},p^s)$ as shown in Fig. 7.

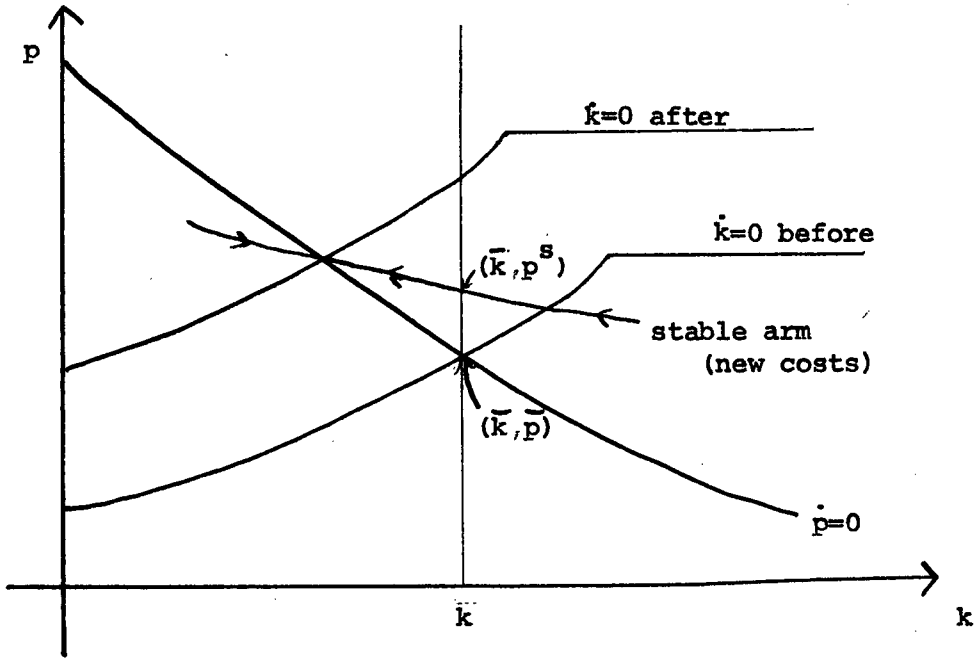
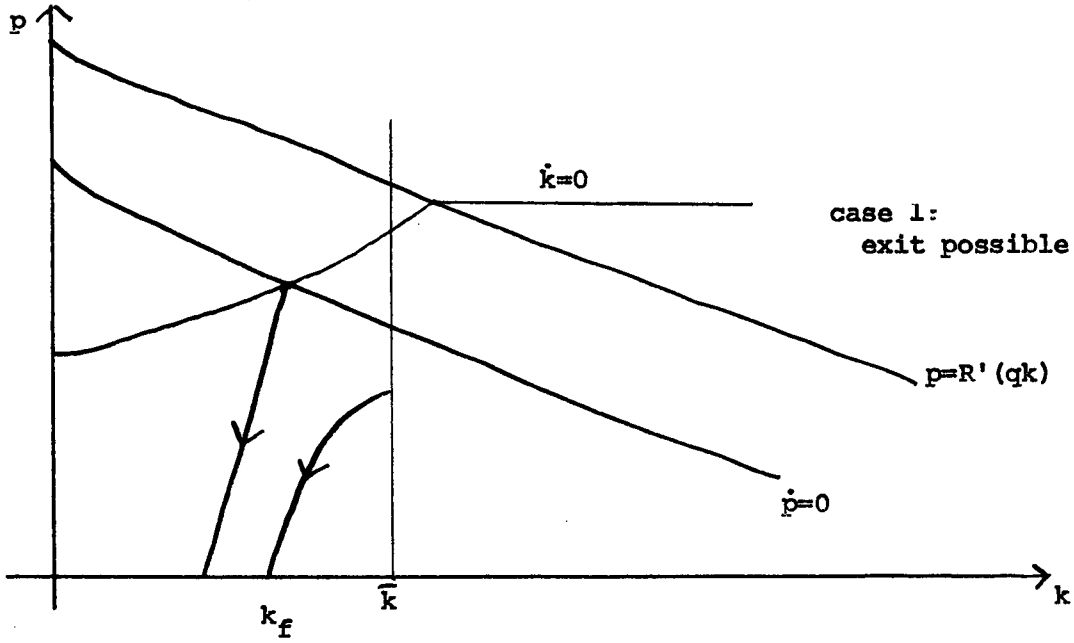


Fig 7.

The "impact effect" of the change can thus be read off at once: shadow price increases from $\bar{p}$ to p^s; stock decreases monotonically from $\bar{k}$ to new equilibrium value; sales (= qk) behave like k; production (= y(p)) drops suddenly but increases as new equilibrium is reached. In this case, then, the long-run comparative statics result that stock will decline is of the same sign as the impact effect, and indeed the effect at all intermediate times.

What if profits are completely eroded and the firm wishes to leave the industry? It may still do so optimally. The transversality conditions (see eq (I-21)) for an optimal exit path are that at the endpoint $M(k_f, p_f, t_f) = 0$ and $p_f k_f = 0$. It is plain from Fig 5 that such endpoints, if one begins from an initial stock holding of $\bar{k}$, can be found only in Region IV, where $M = \{R(qk) - \theta k - C(0) - pqk\} e^{-rt}$. C(0) of course is just fixed cost. Further, since in Region V the equation for the state variable k is $\dot{k} = -qk$, it follows that k will never fall to zero in a finite time,

so that the endpoint must lie on the k-axis. The endpoint, or point of exit, is then given by $p_f = 0$, $R(qk_f) = \theta k_f + C(0)$. It is still not clear that there is any trajectory starting from a point on the line $k = \bar{k}$ that ends at $(k_f, 0)$.



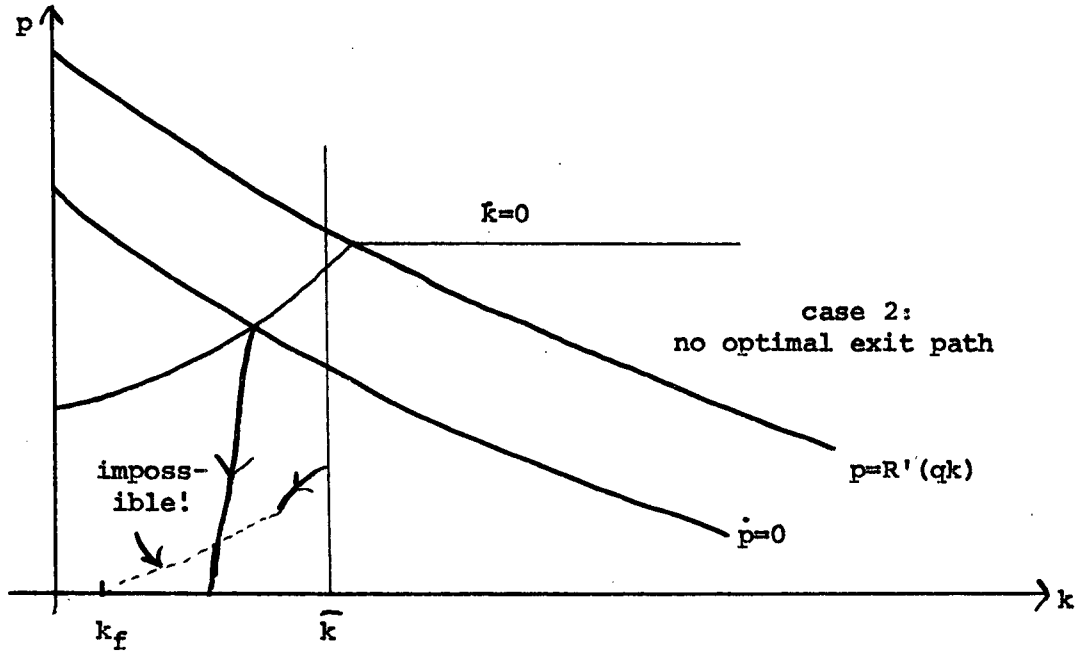


Fig 8.

It can be seen from Fig. 8 that whether or not there is one depends on where the unstable arm reaching into Region IV hits the k-axis. The critical state of affairs is when the unstable arm arrives at the k-axis just at the point k_f. If it is to the left [Case 1] we see that an exit path does exist, if to the right [case 2] then not.

Case 1 is quite easy to understand. There does exist a path at the end of which, i.e., at the point $(k_f, 0)$, the transversality conditions for optimal profit are satisfied. For a firm which does not intend to stay in business, there is no doubt that this is the path that it is best to follow. But in Case 2, although there are paths that end up on the k-axis, for all of them the Hamiltonian there is positive, which implies, according to the discussion of Chapter I, that a path lasting a longer time would be more profitable. It is a familiar result of "turnpike" theory (see, for example, Dorfman, Samuelson and Solow [1958] and Radner [1961]) that the closer an optimal path passes to the saddlepoint, or "turnpike", the longer it lasts. Thus the most profitable of all the motions starting on the $k = \bar{k}$ line is, in Case 2, the stable arm.

Now it may be permissible for the firm to shut down at once without cost, or at some fixed cost. It will of course prefer this course if the best profit available by staying in business either for a finite or for an infinite time is sufficiently negative.

In Case 1, then, the firm has three options: to stay in business along the stable arm, to exit optimally, and to shut down instantly, at cost S, say. In Case 2, there are only two options: the stable arm or instant shutdown. We shall first of all see that in Case 1, optimal exit is always preferred to the stable arm.

The Hamiltonian for this problem involves the time only through an exponential discounting factor, e^{-rt}, and so we know that the marginal worth of time, constant along all optimal paths, is $M - rJ^*$, where J^* is the worth of pursuing the path. The stable arm ends, at time infinity, at the saddlepoint. Because M is multiplied by the factor e^{-rt}, it is plain that $M - rJ^* = 0$ at the saddlepoint at time infinity. Therefore, for the stable-arm path, starting at $(\bar{k},p^s)$ at time 0, the worth is

$$J_s^* = \frac{1}{r} M(\bar{k},p^s,0)$$

$$= \frac{1}{r} (R(q\bar{k}) - \theta\bar{k} - C(y(p^s)) + p^s(y(p^s) - q\bar{k}))$$

[see equation (2)]. At the end of the exit path, $J^* = 0$ by definition and $M = 0$ by the transversality condition. Along this path, too, then, the marginal worth of time is zero. The worth of the path is

$$J_e^* = \frac{1}{r} M(\bar{k},p^e,0)$$

where $(\bar{k},p^e)$ is the beginning of the exit path. But now we observe that $J_e^* > J_s^*$. This follows because $p^e < p^s$, and in Regions IV and V anywhere below the $\dot{k} = 0$ line, $\frac{\partial M}{\partial p} = \dot{k}e^{-rt} < 0$. Thus we confirm that if an optimal exit path is available, it is preferred to the stable arm.

Possible shapes for optimal exit paths are depicted in Fig. 9. All four shapes can be realised in appropriate circumstances. In all cases, stock and sales decline monotonically, but the shadow price may first increase, and it may give rise to continued production for a time, with or without an interruption immediately following the cost change.

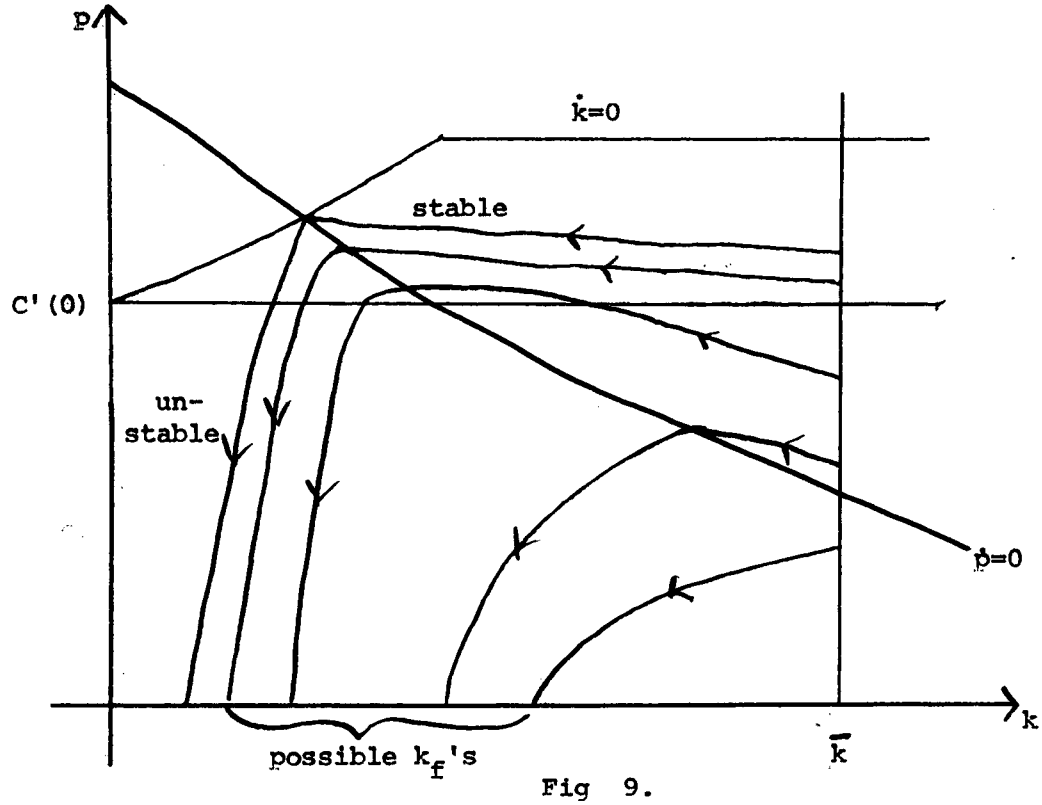


Fig 9.

Case 1 is now fully analysed. Optimal exit by a path shown in Fig 6 is chosen if

$$J^*_e = \frac{1}{r} M(\bar{k}, p^e, 0) > -S$$

and instant shutdown is preferred otherwise. Case 2 is just as easy: the stable arm is followed if

$$J^*_s = \frac{1}{r} M(\bar{k}, p^s, 0) > -S,$$

and otherwise there is instant shutdown.

The "perverse" case of C(y) becoming greater for all y but C'(y) becoming less is no doubt more likely to lead to exit than the usual case discussed above. Most of the possible outcomes are shown in Fig 10 with no further comment.

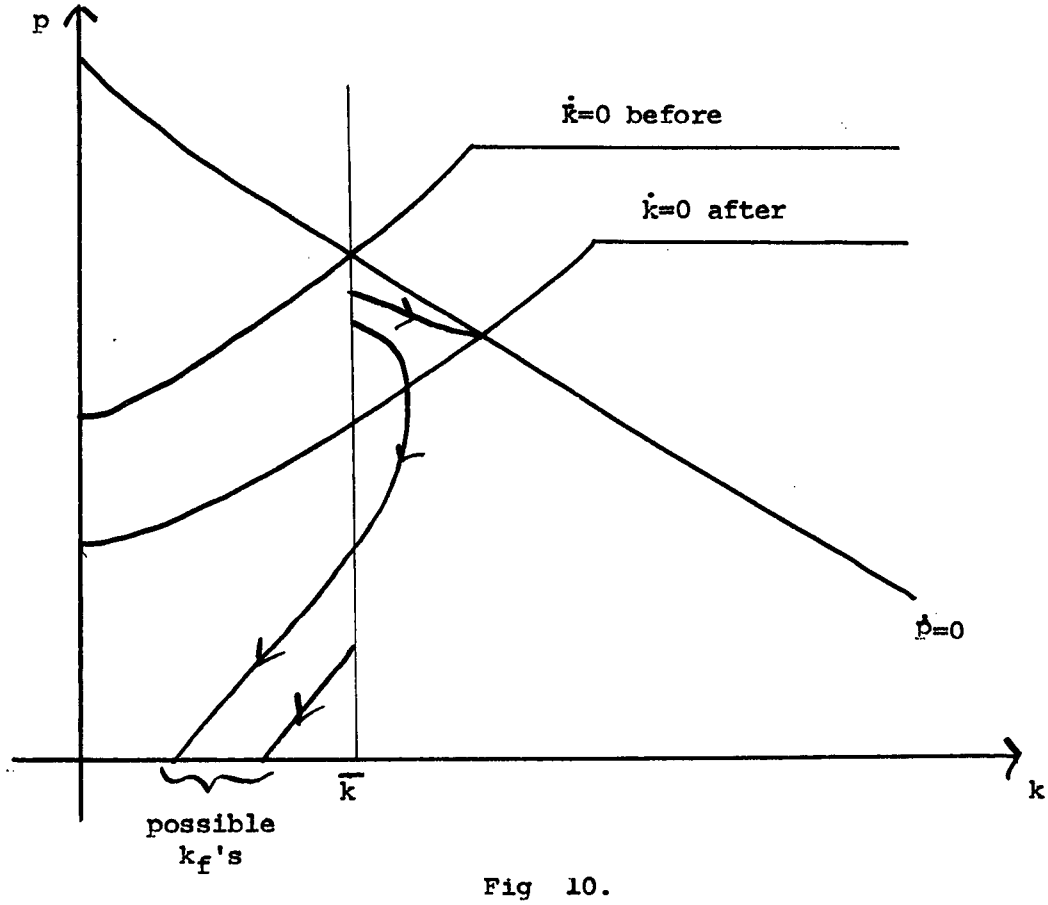


Fig 10.

The intermediate case is where the cost increase is confined to fixed costs, without change in marginal cost. Then the phase plane does not change, and the firm may remain at the saddlepoint with reduced profit if the unstable arm ends to the right of the point $(k_f, 0)$ defined by $H(k_f, 0) = 0$, or exit along the path leading to $(k_f, 0)$ in the other case, or finally shut down at once if that is cheapest.

This model, in its "usual" rather than "perverse" form, is one where

qualitative comparative statics gives an unambiguous answer for the direction of change of k because of a conjugate pair (Samuelson [1947a] in the equations that define the location of the saddlepoint. The impact effect in comparative dynamics is of the same sign as the long-run effect, as are effects at intermediate times (even if exit takes place). One is tempted to believe that the presence of conjugate pairs in equations defining saddlepoints may have stronger (i.e., dynamic) consequences than just the well-known static ones. Against this is the warning conveyed by Fig. 10. This whole matter of impact and long-run effects is discussed at some length by Nagatani [1976], who also draws attention to the extreme difficulties of signing impact effects in problems with more than one state variable. The subject is fascinating, and much remains to be done to elucidate it. The work of Epstein [1977], which treats the Le Chatelier principle in a dynamic context, seems to me to indicate how progress can be made.

In expounding the model used in this essay, I have been precise in specifying the economic meaning of all the variables. I hope that it is clear even so that the same mathematics will describe other problems in other branches of the "theory of the firm." Of particular note is the micro-theory of investment. In this context Lucas [1967] and Gould [1968] have constructed models of intertemporal profit maximisation where adjustment costs arise when a change is made in capital stock, or number of workers employed, or even in rate of investment. Lucas, in particular, has pointed out the need to consider explicitly the matter of entry and exit of firms in an industry when one seeks to explain aggregate investment. For this purpose, as well as for others relating to inventory cycles and the like, the model of

this chapter should be explicitly relevant. But there is a capital-theoretic element involved in almost all of a firm's decisions. The formal similarity of maximising models with stocks of productive capital, inventory, and labour should mean that all of them can be somewhat better understood by means of the techniques used in this essay. In the sense of these remarks, then, the model presented in the next essay is an example which shows how, in one case at least, matters of aggregation and general equilibrium can be handled when entry and exit are taken as endogenous.

CHAPTER IIIA

A MODEL OF URBAN HOUSING.

This chapter and the next make up the last essay of the thesis. Another model of intertemporal maximisation is set up, but this time in a general equilibrium context, with both sides of the market explicitly modelled. It was the claim of Chapter II that interesting comparative dynamics results are likely to be obtainable only if the initial state, on which a perturbing influence is supposed to act, is one of equilibrium. Consequently the present chapter will be devoted to determining the long-run equilibrium of the model, and then in Chapter IIIB a disturbance will be made and its results analysed.

Although the discussion of this chapter cannot strictly be called comparative dynamics, it is, I trust, still of substantial interest in its own right. In order to determine the equilibrium in the model between the forces of demand and supply, even in a steady state, it is necessary to take into account the details of an intertemporal profit maximisation. It will turn out finally that solving the equilibrium equations means locating a fixed point of a mapping, just as in standard general equilibrium theory (see for example Arrow and Hahn (1971)), but here the mapping acts on a function space: it is a highly non-linear integro-differential operator. Extensions of Brouwer's fixed-point theorem apply to function spaces of infinite dimensionality just as well as to the finite-dimensional spaces most commonly used in economic theory however, and so no great new technical difficulty is encountered. The proofs of Brouwer's theorem and its extensions proceed by contradiction and are not constructive, and so these theorems give no help in

finding explicit solutions of equilibrium equations. Another general principle, the contraction mapping principle, is constructive on the other hand, and the possibility of using it is explained at the end of this chapter. In fact, in Chapter IIIB an explicit solution will indeed be found for a rather simplified version of the model of this chapter - the simplification being necessary to make the discussion of dynamics at all tractable. Here, then, the aim is to characterise the equilibrium state by (complicated) equations and to show how existence and uniqueness may be demonstrated in some circumstances.

The model is one of a city inhabited by utility-maximising tenants who live in dwelling-places provided for them by absentee landlords who maximise profits. There are no owner-occupiers. Uncertainty is abstracted from completely, and perhaps a word of justification for this is called for. The overall aim of this thesis is elucidation of some topics in comparative dynamics. Although the effects of uncertainty are, very properly, the object of much study at present, even the comparative statics of a stochastic equilibrium is not yet a solidly-based technique. Consequently, to make any progress in comparative dynamics, I found it impossible to give any attention to uncertainty. I lament this drawback, and hope that economic theory will soon be able to do better. This hope is not motivated only by intellectual curiosity, for central to any assumption that makes an economic agent into an intertemporal maximiser is that he should have well-defined (even if stochastic) expectations about the future over which he is maximising. But the future is always uncertain, even in the presence of futures markets. Throughout this essay, a presumption is made of rational expectations a la Muth (1961). This means simply that economic agents are endowed with perfect foresight of what the model predicts on the assumption that they do have

perfect foresight. If it is accepted that uncertainty is to be ignored, then I feel that a rational expectations hypothesis is the next most honest thing. Besides, it leads to a much cleaner and more self-contained theory than would an alternative hypothesis involving an expectations-generating mechanism leading to frustration and constant planning revision. (See however Goldman (1968) on this subject: such a hypothesis could be made manageable by his kind of scheme.)

Section 1 contains the model of the landlords' behaviour. They have an optimal control problem to solve which is not very different from that of Chapter II. Then in section 2 comes the model of the tenants. Their behaviour is taken to be governed by instantaneous utility maximisation, and the assumption that they can move from one dwelling-place to another costlessly. This scheme, however unrealistic, is usual enough in demand analysis - a consumer's intertemporal considerations are only beginning to be noticed, and would certainly be an unwanted complication here (see Diewert (1974)). In section 3 the long-run equilibrium between landlords and tenants is worked out in the sense discussed above.

(1) <u>The Landlord's Profit</u>

A city of any age contains buildings of widely differing dates of construction. Most often, the older buildings, however solid their structure, are not kept up very well, and provide the not very comfortable, run-down housing of the poor. New buildings on the other hand are regularly fitted out with furnishings of great luxury, and are expected to be inhabited by high-income people. In the model to be discussed in this section, each dwelling will be assumed to be characterised by two properties only: the age of the structure, v, and the level of upkeep, or "comfort", k. This variable, k, is

a stock of upkeep, not a flow. Dwellings are owned by absentee landlords, who invest in upkeep so as to maximise the discounted stream of expected rents they receive. City dwellers are distributed over a spectrum of incomes, according to which, as well as to their tastes, they choose, by maximising their utility, a level of upkeep from the selection offered by the landlords. The term "city dweller" should be understood to mean an entire household rather than an individual, although only one utility function will be allowed to each household. The income of any household will be assumed to be constant over time. City dwellers are presumed to be quite indifferent to the age of the buildings they inhabit - only upkeep is a characteristic entering their utility functions. The municipal authority, on the other hand, cares only about the age of a building. After it has existed for some time, T, it must be torn down to make room for new construction, if new construction is in fact profitable.

If landlords are in a state of perfect competition, each one will perceive a profile of rents available or expected to be available at any time t in return for a level of upkeep k: let this be denoted by the function R(k,t). Any level of upkeep will depreciate unless maintained by investment, and so, if I is the level of investment, one may imagine that upkeep changes over time according to the equation $\dot{k} = I - \delta k$, where δ is the depreciation rate. Let the cost per unit time of level I of investment be C(I). Then each landlord will wish to maximise

$$J \equiv \int_0^{T-v} e^{-\rho t} (R(k,t) - C(I)) \, dt$$

subject to $\dot{k} = I - \delta k$, $I \geq 0$, $k \geq 0$. ρ is the discount rate, v the age of the building. Following as usual the rules of Chapter I, we may form the

Hamiltonian with a current-valued shadow price p:

$$H = e^{-\rho t}(R(k,t) - C(I) + p(I - \delta k)) .$$

The control variable is I, and H is to be maximised with respect to it. The optimal value, I*, is given by:

$$I^* = \begin{cases} I(p) & \text{if } p \geq C'(0) \\ 0 & \text{if } p \leq C'(0) \end{cases} \tag{1}$$

where I(p) satisfies the identity

$$C'(I(p)) = p.$$

The marginal cost function, C′, is assumed, as usual, to be positive and increasing. Further, the function C is taken to be independent of time, which means that technological progress is ignored. In fact this restriction is not very important to the analysis that follows, and could be relaxed at the cost only of complication. The maximised Hamiltonian is

$$M(k,p,t) = e^{-\rho t}(R(k,t) - C(I^*) + p(I^* - \delta k))$$

and so the price equation is

$$\dot{p} = \rho p - R'(k,t) + \delta p \tag{2}$$

where the dash denotes a derivative with respect to k.

The stock equation is of course just

$$\dot{k} = I^* - \delta k \tag{3}$$

and so the phase plane is as shown in Fig 11 .

The equation of the $\dot{p} = 0$ line, the demand curve, is

$$p = \frac{1}{\rho + \delta} R'(k,t)$$

and that of the $\dot{k} = 0$ line, the supply curve, is

$$p = C'(\delta k).$$

Above the line $p = C'(0)$, the stock equation is $\dot{k} = I(p) - \delta k$, and below,

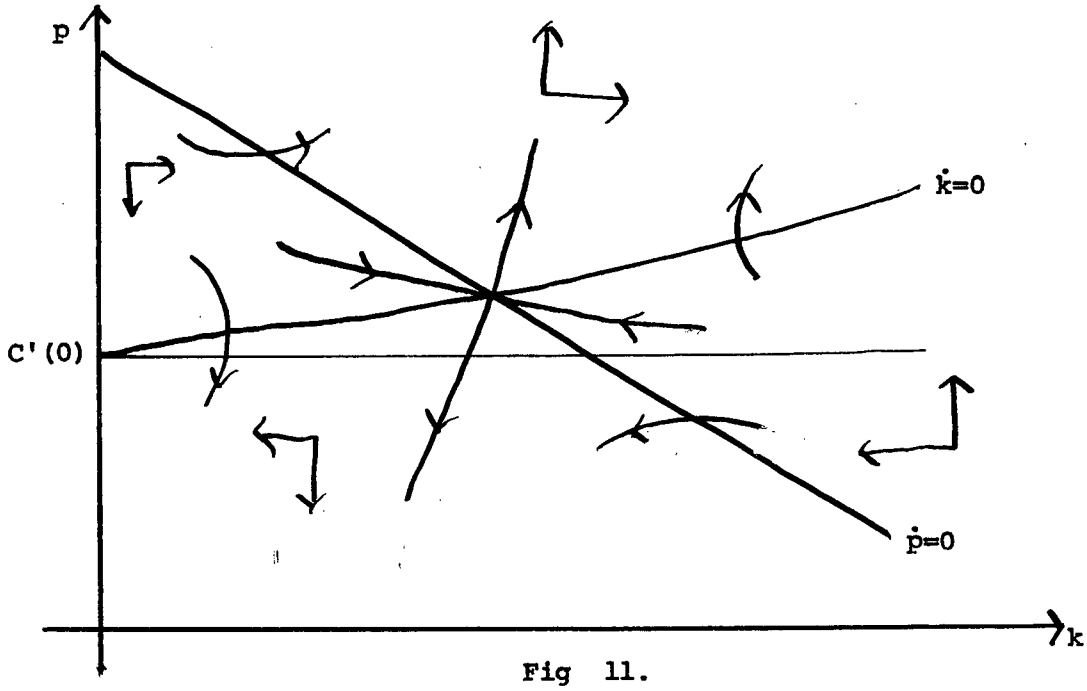

p
k̇=0
C'(0)
ṗ=0
k

Fig 11.

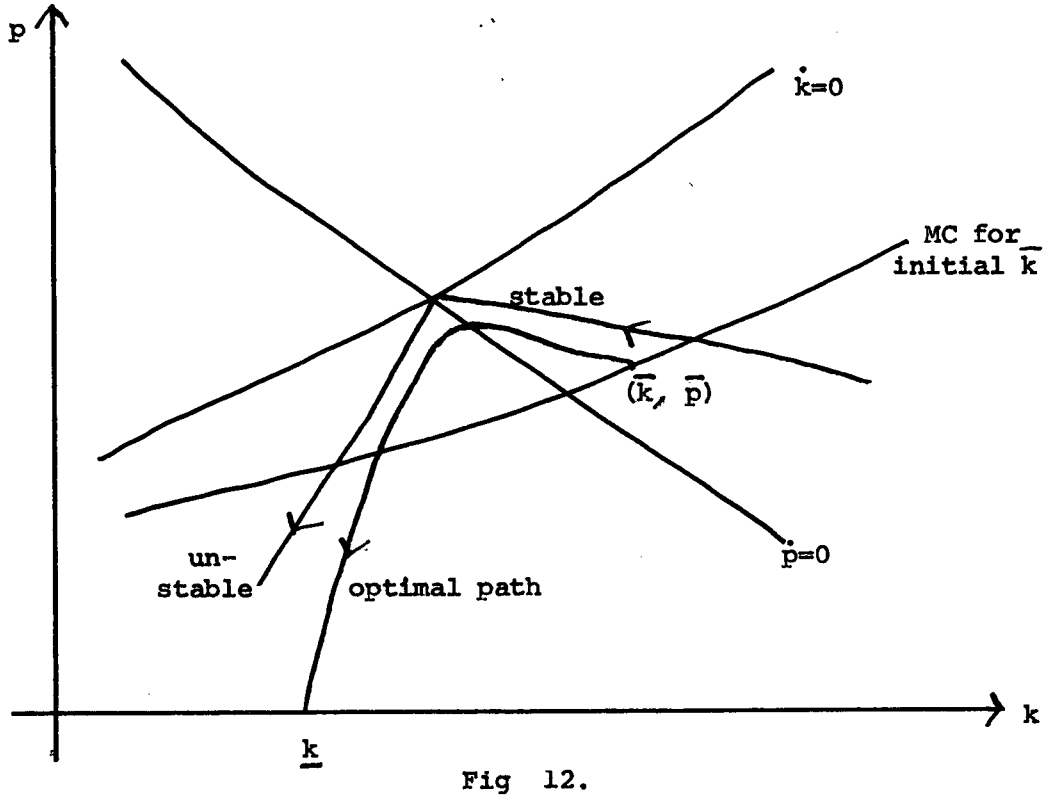

p
k̇=0
MC for
initial k̄
stable
(k̄, p̄)
un-
stable
optimal path
ṗ=0
k
k

Fig 12.

it is just $\dot{k} = -\delta k$.

Because R is an explicit function of t in general, the phase plane is not static. The $\dot{k} = 0$ line does not move, and so the saddlepoint must always be on it, but the $\dot{p} = 0$ line will shift about. This fact introduces no real conceptual difficulties, but it makes the analysis harder. The present chapter, though, is concerned only with a state of long-run equilibrium, in which the time-dependence of R disappears.

Long-run equilibrium will mean zero economic profit for landlords. That is, the worth of optimally exploiting a newly constructed dwelling will be equal to the construction costs.[1] These will be composed of two parts: the building cost, assumed fixed and constant, and the cost of the initial state of comfort, $\bar{k}$, say. Since the line $p = C'(\delta k)$ is the supply curve (marginal cost curve) for comfort levels in an already constructed building, it is reasonable to suppose that costs are lower on the construction site than when tenants are around to object to inconvenience. The marginal cost curve for $\bar{k}$ will thus be supposed to lie *beneath* the $\dot{k} = 0$ line, as shown in Fig 12. The initial point, $(\bar{k}, \bar{p})$, on the optimal trajectory must, by the usual transversality argument, lie on this curve. Similarly, the final point must have zero shadow price for optimality: the optimal trajectory must end on the k-axis. The actual trajectory is then uniquely defined by the total lifetime of the building, T. Paths which go closer to the saddlepoint take longer: the standard turnpike result. The optimal path also defines the final state of upkeep, $\underline{k}$, say. Because the time-dependence is now restricted to the discount factor, the worth of the building can be computed by noting that $M - \rho J^*$ is constant along the path: the result is:

---

1) To avoid any difficulties to do with rent of the land that dwellings are on, it is easiest to assume that land is so abundant that it is a free good.

$$J^* = \frac{1}{\rho}\,(M(\bar{k},\ \bar{p},\ 0) - M(\underline{k},\ 0,\ T)). \tag{4}$$

2. The City Dweller's Choice of a Home.

Each city dweller is endowed with a utility function, U, the arguments of which for our purposes are k, the comfort level of his dwelling, and X, a Hicksian composite of everything else he buys. The price index for the composite is denoted by P. He is assumed to have an unchanging income y: later y will be used to index the inhabitants of the city, and so it will be possible to let different people have different tastes by allowing U to depend on y. We may write then:

$$U = U(k,\ X;\ y).$$

The problem of maximising U is slightly different from the usual one, as k does not measure so much a quantity as a *quality* of dwelling. Each inhabitant is presumed to have a completely inelastic demand for exactly one dwelling. Then his budget constraint is:

$$R(k) + PX = y. \tag{5}$$

The first-order conditions for a utility maximum are:

$$U_k = \lambda R'(k)$$

$$U_X = \lambda P$$

along with equation (5). (λ is a Lagrange multiplier = marginal utility of income.) If P is constant - as it must be in long-run equilibrium - there is no loss of generality in setting it equal to unity. Then the first-order condition may be written thus:

$$R'(k) = \Omega(k,\ R(k),\ y) \tag{6}$$

where Ω is a marginal rate of substitution:

$$\Omega(k,\ R(k),\ y) \equiv \frac{U_k(k,\ y - R(k);\ y)}{U_X(k,\ y - R(k);\ y)}\ .$$

For further analysis all we shall need is this function Ω, at least for this chapter. It should be noted here that, although the utility function depends explicitly on y and thus can in principle accommodate any sort of preferences for each city dweller, it will in the next section be assumed that, for any reasonable rents profile R(k), a higher income y will lead to a choice of a higher k according to eq (6). This assumption, probably quite reasonable, is needed to avoid technical complications.

3. Equilibrium between Landlords and Tenants.

Let us assume that there are N inhabitants of the city, and so in equilibrium N dwellings. Let the income distribution of these tenants be described by a cumulative distribution function F, so that there are NF(y) people with income less than or equal to y. Let $\underline{y}$ be the lowest income, $\bar{y}$ the greatest, so that $F(\underline{y}) = 0$ and $F(\bar{y}) = 1$. Now in the *short run*, the supply of dwellings in a given state of upkeep is completely inelastic. In fact, let the distribution of states of upkeep be represented by a function G, such that the number of dwellings with a comfort level less than or equal to k is NG(k). It is convenient to abstract from all market imperfections and assume that at each moment demand and short-run supply are in instantaneous equilibrium. (In long-run equilibrium, G does not change over time, and this instantaneous equilibrium is identical to the full one.) Once G is given, then R(k), the rents profile, should be completely determined by demand, that is by eq (6).

Each tenant, faced with the profile R(k) and given his income y, can solve eq (6) to determine the comfort level, k, that he will purchase. The result, with the assumption stated at the end of the preceding section, will be a one-to-one increasing relation between y and k - let us write it as

$$y = y(k) \tag{7}$$

with y a function of k rather than *vice versa*.

If the function R is specified, then y(k) is determined by eq (6). Contrariwise, if the function y is specified, the rents profile R(k) can be recovered from eq (6), which is now simply an ordinary differential equation for R, if a boundary condition is available. The desired boundary condition is obtained from the zero-profit condition of long-run equilibrium, which equates the expression in eq (4) with total construction costs. Although it is a long-run condition, it is a legitimate determinant of instantaneous equilibrium because, if new construction is to take place at each moment (and it must since demolition goes on continuously), then the expected rents profile at all times during the life of a dwelling must guarantee zero profit at each instant.

The short-run instantaneous equilibrium, determined purely by demand in the presence of an inelastic supply G(k), can now be written down. Since the number of tenants occupying dwellings of comfort level less than k is NG(k), and this number, by the assumption of the monotonicity of y(k), is just the number of tenants with incomes less than y(k), that is, NF(y(k)), it follows that

$$y(k) = F^{-1}(G(k)) \tag{8}$$

The inverse function F^{-1} is always well-defined because F is a cumulative probability distribution. Now in the short run, the functions F and G are both exogenously given, and so therefore is y by eq (8). The rents profile R(k) can now be obtained as discussed above. Eq (8), with our assumption of instantaneous tâtonnement, holds good also in dynamic situations.

In the long run, it is of course the *supply* of comfort which must adjust fully to obtainable rents. In fact, since the discussion of the landlords' profits shows exactly what time-path upkeep levels will take given a rents profile R(k), we can now compute the supply spectrum G(k), given R(k) in a long-run situation. The result of this computation, along with the short-run one giving R from G, will simultaneously determine both functions and complete our general equilibrium analysis.

We are dealing with a steady state in which all buildings last for a time T and in which the distribution of building ages is rectangular. Thus the number of these which have an age, v, greater than some age $t < T$, is $N(1 - t/T)$. For each age t, there is a unique optimal upkeep, k(t), say, given by the optimal path shown in Fig 12. (Unique because k declines monotonically along the path.) The function k(t) satisfies the optimal equations (2) and (3). The number of dwellings of upkeep level less than k(t) is therefore also $N(1 - t/T)$, that is:

$$G(k(t)) = 1 - t/T. \tag{9}$$

To proceed, then, we must solve eqs (2) and (3) for the function k(t). The solution will be in terms of the rents profile R(k). Each of the optimal equations for the variables k(t) and p(t) can be solved for one variable in terms of the other. From eq (2) we obtain

$$\frac{d}{dt}\left(pe^{-(\rho+\delta)t}\right) = e^{-(\rho+\delta)t}\left(\dot{p} - (\rho + \delta)p\right)$$

$$= e^{-(\rho+\delta)t} R'(k)$$

whence

$$p(t) = \int_t^T dt' \, e^{-(\rho+\delta)(t'-t)} R'(k(t')), \tag{10}$$

since $p(T) = 0$ by the terminal transversality condition.

Similarly from eq (3) we obtain:

$$k(t) = \underline{k}e^{\delta(T-t)} - \int_t^T dt' \, e^{\delta(t'-t)} \, I(p(t')), \tag{11}$$

where $\underline{k}$ (see Fig 12) is the upkeep level at the moment of demolition, and where the function $I(p)$ must be taken as having the value zero for $p \leq C'(0)$ (eq (1)). As can be seen from Fig 12, there is an age of building, t_o say, past which no further maintenance is done, since $p \leq C'(0)$. Therefore eq (11) can be rewritten as:

$$k(t) = \underline{k}e^{\delta(T-t)} - \int_t^{t_o} dt' \, e^{\delta(t'-t)} \, I(p(t')) \quad \text{for } t \leq t_o \tag{12}$$

$$k(t) = \underline{k}e^{\delta(T-t)} \quad \text{for } t \geq t_0 \tag{13}$$

Now eqs (10) and (11) can be combined to yield:

$$k(t) = \underline{k}e^{\delta(T-t)} - \int_t^{t_o} dt' \, e^{\delta(t'-t)} \, I\Big(\int_{t'}^{T} dt'' \, e^{-(\rho+\delta)(t''-t')} \times R'(k(t''))\Big) \quad \text{for } t \leq t_o \tag{14}$$

First of all we observe that eqs (14) and (13) in effect provide the answer to the long-run supply problem in the face of an (exogenous) rents profile $R(k)$. This is so, since if we invert the function $k(t)$, given by eqs (14) and (13) in terms of $R(k)$ and other exogenous functions, we obtain the supply spectrum $G(k)$ from eq (9) :

$$G(k) = 1 - \frac{t(k)}{T}$$

where by definition $t(k)$ and $k(t)$ are inverse functions one of the other:

$$t(k(t)) = t \quad \text{and} \quad k(t(k)) = k.$$

The time t_o, which has so far been defined only in words, is given by the equation

$$C'(0) = \int_{t_o}^{T} dt' \, e^{-(\rho+\delta)(t'-t_o)} \, R'(\underline{k}e^{\delta(T-t')}) \tag{15}$$

(see eqs (10) and (13)).

The boundary value $\underline{k}$ is perhaps best kept as an exogenous parameter at

this point, since its value depends on the construction-site marginal cost schedule for k (Fig 12). For our present purposes, it contains all we need to know about this schedule, since the point $(\bar{k}, \bar{p})$, which lies on it, is just the point (k(0), p(0)) given by eqs (10) and (12) in terms only of $\underline{k}$ and other known quantities.

The next step, then, is to combine eqs (6), (8), (9), (13) and (14), which express the short-run equilibrium and the long-run equilibrium, so as to determine simultaneously both R(k) and G(k). This step is the "general equilibrium" step, and it may be worthwhile to point out the similarities between our equations and the usual general equilibrium ones. It has already been remarked that the short-run equilibrium is characterised by the forces of *demand*. The distribution function G(k), describing as it were the "quantities" of upkeep available, corresponds to a vector of *quantities* in a conventional general equilibrium economy. The R(k) which, for a given G(k), is derived from eqs (6) and (8), then can be seen as a function corresponding to a vector of market-clearing *prices*. The long-run equations determine *supply*; that is, the G(k) derived from eqs (9), (13) and (14) for a given R(k) gives the upkeep levels called forth by the rents R(k). Thus this part of the calculation corresponds to writing down a set of supply functions - a vector of quantities supplied in response to a vector of prices. It can thus be seen that our model is indeed formally analogous to a general equilibrium one, with vectors (finite-dimensional usually) of quantities and dual price vectors replaced by functions in a space of cumulative probability distributions and dual functions in some suitable dual space.

The calculations involved in this general equilibrium step proceed as follows. From eq 14(14), which is the solution of the optimal control equations, some changes of variable lead us to an equivalent equation, (17),

for the inverse function t(k). Then the demand equation, (6), is solved formally for the rents function R(k) in terms of t(k). This allows the equation for t(k), eq (17), to be written in terms of t(k) alone; this will be eq (20). The question of the existence and uniqueness of a solution for this equation is then taken up. First, a series of technical manipulations gives eq (26), which is just eq (20) much simplified. Then a fixed-point theorem is invoked to conclude the analysis.

It is convenient to begin with some manipulations of eq (14). The function t(k) inverse to k(t) is defined, via eq (14), by:

$$k = \underline{k}e^{\delta(T-t(k))} - \int_{t(k)}^{T} dt' \, e^{\delta(t'-t(k))} \times I\Big(\int_{t'}^{T} dt'' \, e^{-(\rho+\delta)(t''-t')} \, R'(k(t''))\Big) \tag{16}$$

where, to save writing two equations at each step, we make the convention $I(p) = 0$ for $p \leq C'(0)$. Eq (13) is now therefore subsumed in eq (14). A change of variable can be performed in each of the integrals in eq (16), from a time-variable to a state-variable. Thus, let $t' = t(k')$; $t'' = t(k'')$. Then it is easy to see that

$$\int_{t'}^{T} dt'' \, e^{-(\rho+\delta)(t''-t')} \, R'(k(t''))$$

$$= \int_{\underline{k}}^{k'} dk'' \Big(-\frac{dt}{dk}(k'')\Big) \, e^{-(\rho+\delta)(t(k'')-t(k'))} \, R'(k''),$$

since $t(\underline{k}) = T$, $t(k') = t'$ and $k(t(k'')) = k''$ by definition. Similarly the integral $\int_{t(k)}^{T} dt' \, e^{\delta(t'-t(k))} \, \Psi(t')$, for any function Ψ of t', can be expressed as: $\int_{\underline{k}}^{k} dk' \left(-\frac{dt}{dk}(k')\right) e^{\delta(t(k')-t(k))} \, \Psi(t(k'))$.

Thus eq (16) becomes, on multiplication by $e^{\delta t(k)}$:

$$ke^{\delta t(k)} = \underline{k}e^{\delta T} - \int_{\underline{k}}^{k} dk' \left(-\frac{dt}{dk}(k')\right)$$

$$\times e^{\delta t(k')} I\left(\int_{\underline{k}}^{k'} dk'' \left(-\frac{dt}{dk}(k'')\right) e^{-(\rho+\delta)(t(k'')-t(k'))} R'(k'')\right). \qquad (17)$$

The above equation holds for all k between $\underline{k}$ and $\bar{k}$.

The demand equation (6) can be written as follows:

$$R'(k) = \Omega(k, R(k), y(k))$$

where the function y(k), as in eq (7), gives the income of the tenant who chooses comfort level k. But from eqs (8) and (9) we know that y(k) can be expressed as:

$$y(k) = F^{-1}(G(k))$$
$$= F^{-1}\left(1-\frac{t(k)}{T}\right)$$

with the same function t(k) as in eq (17). Therefore we obtain the following ordinary differential equation for R(k):

$$R'(k) = \Omega\left(k, R(k), F^{-1}\left(1-\frac{t(k)}{T}\right)\right). \qquad (18)$$

For a unique solution, a boundary condition is needed. We shall simply take $R(\underline{k})$ as $\underline{R}$, and finally pin down $\underline{R}$ by the zero-profit condition. Meanwhile $\underline{R}$ will be treated as another parameter of the model. In fact, just as a knowledge of $\underline{k}$ was enough information about the construction-site marginal cost schedule for present purposes, so is $\underline{R}$ enough information about fixed construction costs. If $\underline{R}$ is given, then we may *infer*, by reasoning backwards, what the cost of constructing a new dwelling must be. It is as legitimate to regard $\underline{R}$ as truly exogenous as it would be to introduce a constant C, say, as the cost of a new dwelling with built-in comfort level $\underline{k}$, and then solving for $\underline{R}$ - and it is very much simpler.

A standard Lipschitz condition (see, as a convenient if not very complete reference, Bryson and Ho (1969a)) is then enough to guarantee the existence and uniqueness of the solution to eq (18). The assumption that such a condition is satisfied is not at all stringent, and is hereby made. The solution can then be formally written as:

$$R(k) = \Delta(t(k);\ \underline{R}) \tag{19}$$

where $\Delta(\ldots;\ \underline{R})$ is some, in general non-linear, operator acting on the space of continuous functions t from the interval $(\underline{k},\ \bar{k})$ into the non-negative real line. (Or, if we wish to be more precise, although it would seem to be of little benefit, functions such that $t(\underline{k}) = T$, $t(\bar{k}) = 0$.)

Let $\Delta'(t(k);\ \underline{R})$ denote the function of k which is the derivative with respect to k of $\Delta(t(k);\ \underline{R})$. Then eq (17) becomes:

$$ke^{\delta t(k)} = \underline{k}e^{\delta T} - \int_{\underline{k}}^{k} dk' \left(-\frac{dt}{dk}(k')\right) e^{\delta t(k')}$$

$$\times\ I\left(\int_{\underline{k}}^{k'} dk'' \left(-\frac{dt}{dk}(k'')\right) e^{-(\rho+\delta)(t(k'')-t(k'))}\Delta'(t(k'');\ \underline{R})\right) \tag{20}$$

This is now a non-linear integro-differential equation for the function t(k) written in terms of exogenously given quantities only. Once we are satisfied that it possesses a unique sensible solution, our problem is solved, since G(k) is just $1 - \frac{t(k)}{T}$, and R(k) is just $\Delta(t(k);\ \underline{R})$. All that remains to be shown, then, is this matter of existence and uniqueness.

Let us make the definition:

$$\tau(k) \equiv e^{\delta t(k)}. \tag{21}$$

Then $\tau'(k)$ - the dash denotes the derivative - is equal to $\delta\frac{dt}{dk}(k)e^{\delta t(k)}$, and eq (20) can be written in terms of $\tau(k)$ as follows:

$$k\tau(k) = \underline{k}e^{\delta T} - \frac{1}{\delta}\int_{\underline{k}}^{k} dk' \; (-\tau'(k'))$$

$$\times \; I\left(\{\tau(k')\}^{\frac{\rho+\delta}{\delta}} \; \frac{1}{\delta}\int_{\underline{k}}^{k'} dk'' \; (-\tau'(k'')) \; \{\tau(k'')\}^{-\frac{\rho+2\delta}{\delta}} \; \Gamma(\tau(k''); \; \underline{R})\right) \tag{22}$$

where $\Gamma(\tau(k''); \; \underline{R}) = \Delta'\left(\frac{1}{\delta}\log \tau(k''); \; \underline{R}\right)$ and is just another nonlinear operator.

Now $\quad k\tau(k) - \underline{k}e^{\delta T} = \int_{\underline{k}}^{k} \frac{d}{dk'}(k'\tau(k'))\, dk' \qquad$ (since $\tau(\underline{k}) = e^{\delta T}$)

$$= \int_{\underline{k}}^{k} (\tau(k\ ) + k'\tau'(k'))\, dk'.$$

It follows that eq (22) can be written

$$\int_{\underline{k}}^{k} dk' \; \{\tau(k') + \tau'(k') \; \left(k' - \frac{1}{\delta}\, T(\tau(k'); \; \underline{R})\right) \} = 0 \tag{23}$$

with the non-linear operator T defined by:

$$T(\tau(k'); \; \underline{R}) =$$

$$I\left(\{\tau(k')\}^{\frac{\rho+\delta}{\delta}} \; \frac{1}{\delta}\int_{\underline{k}}^{k'} dk'' \; (-\tau'(k'')) \; \{\tau(k'')\}^{-\frac{\rho+2\delta}{\delta}} \; \Gamma(\tau(k''); \; \underline{R})\right). \tag{24}$$

The integral sign in eq (23) is plainly unnecessary:

$$-\frac{\tau'(k)}{\tau(k)} = \frac{1}{k - \frac{1}{\delta}\, T(\tau(k); \; \underline{R})}$$

This equation will be re-integrated in a moment. Meanwhile let us take note of some restrictions that we wish to place on admissible functions $\tau(k)$. In addition to the boundary condition $\tau(\underline{k}) = e^{\delta T}$, we require that $\tau(k)$ is always greater than unity and decreasing, because of the physical interpretation of $t(k)$. Again, since $I(p) = 0$ for $p \leq C'(0)$, it is necessary that, for k less than some k_o, we have simply:

$$\tau(k) = (\underline{k}/k)\, e^{\delta T} \text{ , i.e., } t(k) = T - \frac{1}{\delta} \log (k/\underline{k}) \text{ .}$$

Whatever the behaviour of $\tau(k)$ for $k > k_o$, it follows from eq (18) and its solution eq (19) that, for $k \leq k_o$,

$$\Gamma(\tau(k);\ \underline{R}) = \Delta'(t(k);\ \underline{R}) = \Delta'(T - \frac{1}{\delta} \log (k/\underline{k});\ \underline{R}) \text{ .}$$

If this last result is substituted into eq (24), k_o is determined as the greatest k for which $T(\tau(k);\ \underline{R})$ is zero - that is, k_o is determined completely by the exogenous quantities, as are also $\tau(k_o)$ and $t(k_o)$.

Bearing in mind the definition of $\tau(k)$, eq (21), we may now integrate eq (25) from k_o to k. The result is:

$$t(k) = t(k_o) - \int_{k_o}^{k} dk' \frac{1}{\delta k' - T(t(k');\ \underline{R})} \tag{26}$$

with the operator T now redefined in an obvious manner to act on t(k) instead of $\tau(k)$.

Eq (26) is a rather straightforward looking non-linear equation, and we may ask directly about the existence and uniqueness of its solution. A solution is plainly a fixed point of the operator

$$A(t(k)) \equiv t(k_o) - \int_{k_o}^{k} dk' \frac{1}{\delta k' - T(t(k');\ \underline{R})}$$

It should be pointed out here that $\bar{k}$, the starting-value of the state variable on the optimal trajectory that we are looking for, is still endogenous. It is determined by the equation $t(\bar{k}) = 0$. But we shall never be interested in values of t(k) for $k > \bar{k}$, and so eq (26) could be rewritten as:

$$t(k) = \max\ (A(t(k)),\ 0).$$

Let us trivially redefine the operator A to be the right-hand side of this equation. Then it follows that the denominator $\delta k' - T(t(k'); \underline{R})$, which is positive ($= \delta k_o$) at $k' = k_o$, must always be positive for $k' < \bar{k}$, for otherwise it would vanish and make $t(k)$ equal to minus infinity. We conclude then that A maps positive decreasing functions $t(k)$ $(k_o \leq k < \infty)$ into positive decreasing functions. The boundary value $t(k_o)$ is preserved, and the operator is clearly bounded.

The *contraction mapping principle* (see Krasnosel'skii (1964) and Kleider *et al.* (1968)) can be invoked in cases such as the present to prove existence and uniqueness of a fixed point. Bounded continuous functions defined on (k_o, ∞) form a Banach space with norm defined by:

$$\| t \| = \sup_{k_o \leq k < \infty} | t(k) | \tag{27}$$

The contraction mapping principle says that if A maps a bounded region of this space (such as positive decreasing functions with $\| t \| \leq t(k_o)$) into itself, and if, for any two functions $t_1(k)$ and $t_2(k)$ belonging to this region, the Lipschitz condition

$$\| At_1 - At_2 \| \leq \alpha \| t_1 - t_2 \| \tag{28}$$

holds with $\alpha < 1$, then there exists a unique fixed point of the operator A in the region, which can be calculated from any starting function $t_o(k)$ by iterations:

$$t_n(k) = At_{n-1}(k).$$

We need only ask, then, under what conditions the inequality (28) is satisfied. The iterations can of course be thought of as steps in some tâtonnement process. In fact, many cases in which Brouwer's theorem is used can be so interpreted, and often the contraction mapping principle would be applicable to these cases. If so, because of its constructive

nature, it is greatly to be preferred. One instance of the use of the contraction mapping principle in economic theory can be found in Brock (1972), where it is used in connection with equilibrium forecasting.

It is clear from eq (26) that A is a continuous operator when the norm of eq (27) is used. I have not succeeded in demonstrating that the inequality (28) is always satisfied for any exogenous functions Ω, F and I, but this is not, I feel, a very urgent matter. The contraction mapping principle is a very stringent sufficient condition for existence and uniqueness, and when its requirements are not satisfied, there is often no difficulty in proving at least existence by other fixed-point theorems, many of which are presented in Krasnosel'skii's book. One may quite easily sit down and find out if inequality (28) is or is not satisfied if Ω, F and I are given. When, in the next chapter, specific choices of these functions are made, this will be done, and the solution will in fact be explicitly derived.

The specification of the long-run equilibrium is now finished. Because it is necessarily a steady atate, the solution functions G(k) and R(k) do not depend on time. This means that the equilibrium is rather insensitive to the expectations-generating mechanism postulated for the landlords. As well as the rational expectations assumed at the beginning of this chapter, expectations of rents remaining as they are at any moment would give the same result. In the dynamic analysis of Chapter IIIB however, expectations become more important, and whatever results are found depend on how they are generated. The conditions for "entry and exit", too, are simple because of the steady-state nature of the solution. Again, these conditions assume greater interest in a dynamic context.

CHAPTER IIIB.

A MODEL OF URBAN DECAY.

The long-run equilibrium treated in Chapter IIIA is the starting point for this chapter. Into the steady state of that equilibrium will come a disturbance. Specifically, it will be imagined that a consumption externality arises, adversely affecting housing, but without other effects. This is readily modelled by replacing the utility function of the tenants by

$$U(u(k, \alpha), X; y) \tag{29}$$

where U is the same utility function as before, but where the worth to a tenant of a level of comfort k is no longer simply k but is measured by the function $u(k, \alpha)$. The parameter α measures the externality: we assume that $u(k, 1) = k$, and that $u_\alpha > 0$. The separability of k and α from the other variables in the argument of U expresses the assumption that the externality affects only housing. The specification that $u(k, 1) = k$ means that the value of unity for α corresponds to the absence of any externality: if, as we suppose, the externality that arises is adverse, it must, since $u_\alpha > 0$, correspond to a value of α less than unity. For example, if pollution of some kind, from a smoky factory chimney for instance, is the source of the externality, then clean air means that $\alpha = 1$, and α can be thought of as the inverse of some measure of concentration in the air of noxious fumes.

The spirit of comparative statics would have us now consider the derivatives with respect to α, evaluated at $\alpha = 1$, of all the endogenous

variables of the model. We shall not do so, but rather look at the results of a sudden discrete change from $\alpha = 1$ to a value of α less than 1. Again, comparative statics would focus attention on the functions R(k) and G(k) that measure rents and supply of comfort, and would take account only implicitly of the possibilities of entry and exit from the landlord business (by taking a zero-profit condition on new construction as one of the equilibrium equations). But here, entry and exit are considered quite explicitly, and windfall gains or losses become a feature of our model. These disequilibrium phenomena in fact give the distinctive colour to the economic tale to be told.

For, once the externality has arisen, the bright idea will no doubt occur to someone that the nuisance of building a house in the suburbs and commuting daily is outweighed by the advantage of escaping the externality, and so he moves out of the city. It will presumably be the richest person who moves, since his desired standard of comfort, $\bar{k}$, has become effectively unobtainable. But his departure lowers the demand for housing, and therefore also the revenue profile, R(k), perceived by landlords. If, as it is reasonable to suppose, the *marginal* revenue profile, $R'(k)$, is also lowered, it can be seen from Fig 12 that the landlords will respond by spending less on upkeep. This further lowers the available standards of comfort, and so, in a cumulative process, more and more city-dwellers are pushed to the margin where it is more advantageous to move to the suburbs. Suburban housing will no doubt take time to build, and will be subject to increasing costs. If so, ultimately a cut-off point may be reached where poor people are left in the city, trapped there in decaying slums no longer kept up by the landlords, but without enough income to

afford the now expensive suburban housing.

The above discussion provides a possible scenario of events. If it is to be modelled, then it is clear at once that the conditions for entry of a firm into the suburban housing business and for exit from (and possible re-entry into) the urban construction and renting business must be explicitly laid down.

In section 1, then, a model of the suburbs is appended to our existing model of the city, and entry and exit are discussed. Then in section 2, the assumption is made of a horizontal marginal cost schedule for upkeep in the city. This permits a great simplification of the urban model, and the explicit form of the long-run equilibrium is worked out. Other specific functional forms for exogenous quantities are chosen here and in later sections so as to make it possible to write down explicit expressions for some of the endogenous functions. In section 3, it will be seen that two distinct states of affairs can arise in the coupled city/suburb model, according to whether or not the rate of new suburban construction is fast enough to leave unoccupied dwellings standing in the city. The case in which it is is analysed in section 3, and the case in which it is not in section 4. After this, it is possible in section 5 to catalogue the various modes of urban decay that can be generated by the present model.

1. The Flight to the Suburbs.

The model proposed here for suburban affairs is simpler than the city model. In particular, intertemporal considerations for suburban landlords, tenants and/or owner-occupiers will be abstracted from. We shall in fact lose interest in a person more or less at the moment he or she moves to the suburbs. The only quantities of interest will be the

rental price (for one unit of time) that must be paid for a suburban dwelling of comfort κ at a time when M dwellings exist in the suburbs, and the cost (expressed as a rental price) of constructing such a dwelling. This information is enough to allow calculation of the margin between city and suburb, and, provided only that suburbanites stay in the suburbs once they get there, knowledge of affairs at the margin is all that we shall need.

A very elementary sort of geography is implied in all this. Both the city and the suburb are treated as points, or as completely undifferentiated areas. People may make only a binary choice of where to live - notions of better or worse neighbourhoods are ignored.

Our programme for this section is as follows. First we shall specify the response by the construction firms that build suburban housing to any given state of demand. In particular entry and exit of these firms will be discussed. Then demand considerations will be taken up, and it will be seen that demand for suburban dwellings is determined by the margin of indifference between city and suburban living, this margin being determined by conditions in the city. Lastly the two sides of the market will be brought together to give the dynamics of suburban construction.

Let us then write $S(\kappa, M)$ for the rent of a suburban dwelling of comfort level κ at the time when exactly M suburban dwellings exist, and $\Xi(\kappa, M)$ for the cost of constructing it, expressed as a rental charge. The profit obtained by a suburban construction firm for putting up this dwelling is thus $S(\kappa, M) - \Xi(\kappa, M)$ per unit time, or, more sensibly, the capitalised sum $\frac{1}{\rho}(S(\kappa, M) - \Xi(\kappa, M))$, where ρ is the discount rate.

Next, we must enquire how many firms are engaged at any time in suburban construction, how fast they can put up a dwelling, and what state of competition prevails among them. For simplicity, we may assume that each firm can put up exactly one dwelling per unit time, regardless of its comfort level, κ. The construction rate at any moment is then just $m(t)$, where the function $m(t)$ gives the number of firms in the business at time t. These "firms" are to be thought of as entrepreneurial units in competition, and so it seems reasonable to require that the profit that each makes per unit time should be the same across firms at each moment. Why, in a state of competition, should there be any profit at all? After all, in the city, a zero-profit assumption was used to characterise equilibrium. Various answers can be given: adjustment costs of one kind or another, rezoning costs, or, in general, costs of entry. If, for instance, there are many potential entrants into the suburban construction business, and each perceives a barrier to entry (in dollar terms) of a different size, then, the more firms are in existence, the higher profits must be - here "profit" means earnings over and above the prime cost of all used-up factors of production. At all events, profit may certainly exist in a competitive industry where entry and exit are not instantaneous adjustment processes, even if no dollar cost is involved. But once firms do exist and are in competition, it is a reasonable assumption that the firms will so bid for business that every dwelling under construction at a given moment, no matter what its comfort level κ, will yield the same profit. Otherwise, low-profit firms would constantly have an incentive to underbid high-profit firms.

If the above reasoning is accepted, then we may assume that, if *exit* of a firm is costless, or rather, if, whatever exit costs may be, each firm can be imagined to exit after each dwelling that it is building is completed and subsequently to re-enter if the profit rate is satisfactory, then, the number of firms, $m(t)$, is an increasing function of the profit per dwelling alone, that is:

$$m(t) = f(\pi(t)), \qquad (f' > 0) \tag{30}$$

where $$\pi(t) = S(\kappa, M) - \Xi(\kappa, M) \tag{31}$$

is the profit. Our assumptions mean that $S(\kappa, M) - \Xi(\kappa, M)$ must be *independent* of κ over the existing spectrum of comfort levels κ, so that $\pi(t)$ is well defined.

The cost function $\Xi(\kappa, M)$ is of course exogenously given, and we shall assume that $\Xi_\kappa > 0$, $\Xi_M > 0$, the second of these conditions expressing increasing costs in suburban construction as more dwellings are put up. A justification of this is no more than an invoking of the law of diminishing returns with some factor (available drained land or some such) fixed.

The state of affairs in the suburbs is determined by $M(t)$, the number of finished dwellings, and we have by definition that:

$$\dot{M}(t) = m(t) = f(\pi(t)) \tag{32}$$

Since $M(0) = 0$ (the externality begins at time zero), eq (32) gives the dynamics of the suburb once $\pi(t)$ has been fully expressed in terms of the exogenous variables. This means that we must now pin down $S(\kappa, M)$, presumably by the forces of demand. Potential suburbanites have the alternative of living in the city, and so we may begin our demand analysis there. A city-dweller of income y has, by eq (29), utility:

$$U = U(u(k, \alpha), y - R(k); y).$$

($R(k)$ is, as usual, the rents profile in the city at some moment - tenants have no intertemporal considerations.) The first-order condition for maximising this utility is

$$R'(k) = \frac{u_k(k, \alpha)U_u(u(k, \alpha), y-R(k); y)}{U_x(u(k, \alpha), y-R(k); y)}$$

$$= u_k(k, \alpha)\Omega(u(k, \alpha), R(k), y) \tag{33}$$

where Ω is the marginal rate of substitution introduced in eq (6). Let the solution, k, of this first-order condition be written $k = k(y)$.

In general there is some disadvantage associated with living in the suburbs rather than in the city (else why was there no suburb before the externality?) and it is convenient to model this fact by assuming that a standard of comfort in the suburbs κ enters utility functions as $u(\kappa, \alpha_s)$ for some α_s satisfying $1 > \alpha_s > \alpha$. This number α_s does *not* of course correspond to an externality, but rather an intrinsic disadvantage of suburban life. Then, analogously to eq (33), we obtain for the suburbs:

$$S'(\kappa) = u_k(\kappa, \alpha_s)\Omega(u(\kappa, \alpha_s), S(\kappa), y) \tag{34}$$

Let the solution of this equation be $\kappa = \kappa(y)$. (Time-dependence, as expressed via either t or M, is suppressed for the moment for the sake of clearer notation.)

Then, for a person of income y to be at the margin of indifference between city and suburb, we must have that:

$$U(u(k(y), \alpha), y-R(k(y)); y) = U(u(\kappa(y), \alpha_s), y-S(\kappa(y)); y). \tag{35}$$

From this equation and eq (34) we can now determine $\pi(t)$. From eq (31) it follows that $S'(\kappa, M) = \Xi'(\kappa, M)$ (dashes here denote differentiation with respect to κ), so that eq (34) becomes:

$$\Xi'(\kappa, M) = u_k(\kappa, \alpha_s)\Omega(u(\kappa, \alpha_s), \pi(t) + \Xi(\kappa, M), y). \tag{36}$$

Now, if at time t, there are M dwellings in the suburbs, then there are also M people in the suburbs, out of a total of N people altogether, and so the income of the person at the margin, y(M) say, must satisfy the equation

$$N(1-F(y(M))) = M, \text{ i.e. } y(M) = F^{-1}\left(\frac{N-M}{N}\right). \tag{37}$$

Strictly speaking, for this to be true we must make an assumption like that of section 2 of Chapter IIIA that people leave the city in descending order of income. For any specific choices of the exogenous functions, this is a matter to be verified, not assumed. But for the moment, the assumption is all that is necessary. If, then, $F^{-1}\left(\frac{N-M}{N}\right)$ is substituted for y into eq (36), the value of κ which satisfies the equation, $\kappa(M, \pi(t))$, say, (time-dependence is explicit again) is the level of suburban comfort chosen by the person at the margin. Consequently we may use eq (35) to obtain:

$$U(u(k(y(M)), \alpha), y(M) - R(k(y(M)), t); y(M))$$
$$= U(u(\kappa(M, \pi(t)), \alpha_s), y(M) - \pi(t) - \Xi(\kappa(M, \pi(t)), M); y(M)) \tag{38}$$

The left-hand side of this equation involves, in addition to M and t, only the exogenous functions and quantities, U, u, α, y(M) (via F), and the functions k(y) and R(k, t) which depend only on the state of affairs in the city. For the purposes of the model of the suburbs, these last are taken as given, and so the left-hand side can be regarded as a known function of M and t. But, on the right-hand side, U, u, y(M), α_s, Ξ and the function $\kappa(M, \pi(t))$ are all exogenous or derived directly from exogenous quantities. The result is eq(38) is an equation which can be solved for $\pi(t)$ as a function of M and t alone. When this solution is substituted into eq (32), the dynamics of the suburban model have been fully specified.

2. A Particular Case of the Model of the City.

In this section, the marginal cost function $C'(I)$ for investment in upkeep in city dwellings will be assumed to be a constant, c, for values of I between zero and $\delta\bar{k}$, and for higher values of I infinite. The aim of this assumption is to make especially simple the optimal upkeep path that city landlords will follow. That it does so can be seen by observing that the function $I(p)$ of eq (1) becomes equal to the constant $\delta\bar{k}$ if $p > c$, 0 if $p < c$, and indeterminate between these values for $p = c$. This is in fact an instance of a "bang-bang" control (see Bryson and Ho (1969b)). The result is that if the state variable k has the value $\bar{k}$, it will remain unchanged at that value for as long as $p > c$. If the further assumption is made that the construction-site marginal cost schedule is given by the *same* function $C'(\delta k)$ as gives the $\dot{k} = 0$ line, then, so long as a dwelling lasts long enough to receive some positive amount of maintenance, it will be put up originally with comfort level $\bar{k}$ and maintained there until, at the end of its life, it decays according to the equation $\dot{k} = -\delta k$.

In Fig 13, the phase plane, analogous to that of Fig 12, is drawn for this case. The exact location of the $\dot{p} = 0$ line, with equation

$$p = \frac{1}{\rho + \delta} R'(k, t)$$

(dash denotes differentiation with respect to k), affects only the details of the "exit path", so long as $(1/(\rho + \delta))R'(\bar{k}, t) \geq c$ for all t (39)
As usual, if specific choices of the exogenous functions and quantities of the model are made, it is necessary to verify that this inequality is satisfied.

To proceed, then, we shall first compute, for the steady state as

described in Chapter IIIA, the supply function G(k) and then the rents function R(k). On the way, some specific choices will be made for some of the exogenous functions. Then various checks willbe made to ensure the consistency of the solution with the various assumptions that have been made. In the course of these checks, it will turn out that the model can be understood rather more generally than has been stated so far, and this will be explained. Finally we shall see that eq (26) of Chapter IIIA gives the same solution as the one obtained here.

We begin by observing that, in a steady state with a rectangular distribution of building ages, the comfort supply function G(k), is given by: $G(k(t)) = 1 - (t/T)$ (eq (9))

while

$$k(t) = \begin{cases} \bar{k} & (t \leq t^*) \\ \bar{k}e^{-\delta(t-t^*)} & (t \geq t^*) \end{cases} \tag{40}$$

where t* is the (still-to-be-determined) building age at which maintenance stops. Thus:

$$G(k) = 1 - \frac{1}{T}\left(t^* + \frac{1}{\delta}\log\left(\frac{\bar{k}}{k}\right)\right) \quad \text{for } k < \bar{k} \qquad \Big\} \tag{41}$$

and $G(\bar{k}) = 1.$

We may notice further that

$$\underline{k} = \bar{k}e^{-\delta(T-t^*)} \tag{42}$$

($\underline{k}$, as usual, denotes the upkeep level at the time of demolition.)

With the help of eq (41) we may deduce from eq (8) the function y(k) to be used in eq (6) in order to determine R(k). We have:

$$y(k) = F^{-1}(G(k)) = F^{-1}\left(1 - \frac{1}{T}\left(t^* + \frac{1}{\delta}\log(\bar{k}/k)\right)\right) \quad (k < \bar{k}) \tag{43}$$

and $y(\bar{k})$ is now the *set* of all incomes above $F^{-1}(1 - (t^*/T))$.

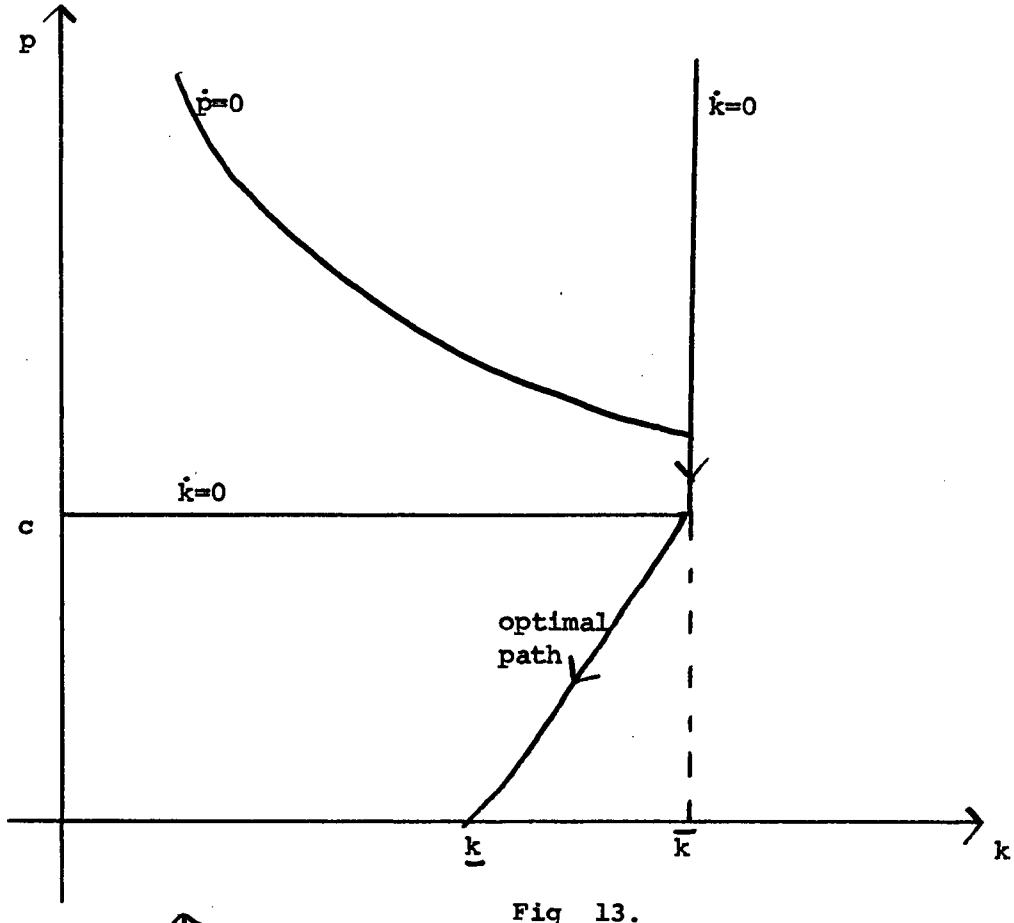


Fig 13.

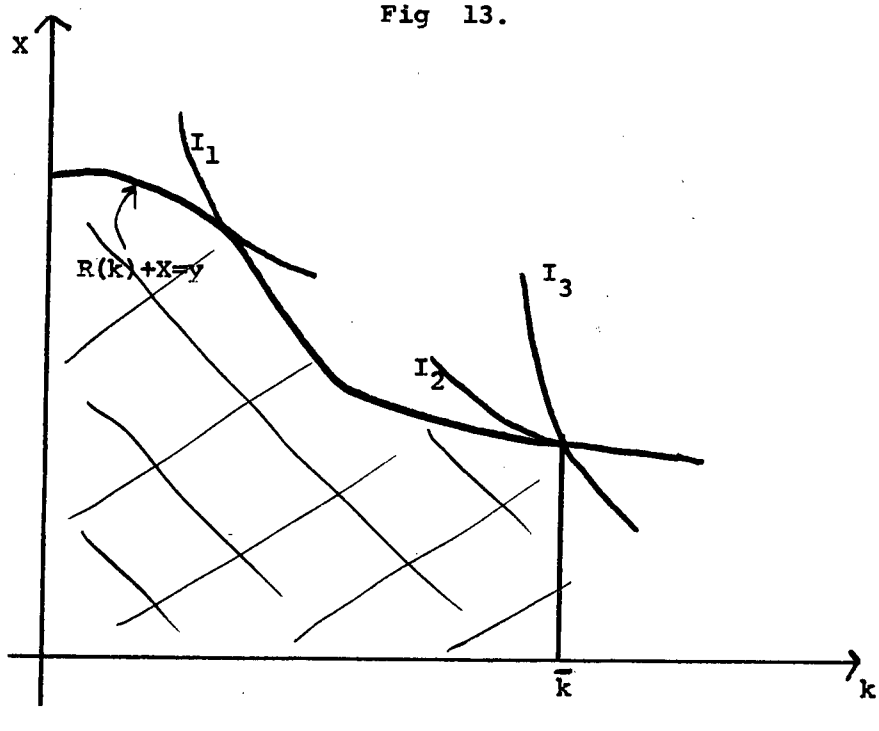


Fig 14.

In what way can this transformation of the function $y(k)$ into a correspondence be justified? Since it is impossible, because of the infinite marginal cost, for a comfort level exceeding $\bar{k}$ to exist, we may write symbolically $R(k) = \infty$ for $k > \bar{k}$. The budget set of a tenant of income y, given by points (k, X) satisfying the inequality $R(k) + X \leq y$, is then truncated by a vertical line at $k = \bar{k}$ (see Fig 14, in which for clarity the budget lines of different households have been rescaled so as to coincide.) Tenants whose income is less than $F^{-1}(1 - (t^*/T))$ will be on indifference curves like I_1 tangent in the usual way to the boundary of the budget set (cross-hatched). Tenants whose income exceeds $F^{-1}(1 - (t^*/T))$ will be on indifference curves like I_3, with no tangency at the corner of the budget set. The person whose income is exactly $F^{-1}(1 - (t^*/T))$ will be on indifference curve I_2, tangent to the non-vertical part of the boundary of the budget set just at the corner. All this means that eq (6) gives $R(k)$ just as before for $\underline{k} \leq k \leq \bar{k}$ if $y(\bar{k})$ is interpreted simply as $F^{-1}(1 - (t^*/T))$.

This is an appropriate time to introduce some more specific choices of exogenous functions. We shall be able, after doing so, to perform all the verifications necessary to ensure that the model makes sense. Accordingly, let us set:

$$F^{-1}(x) = \underline{y} + bx \qquad (0 \leq x \leq 1) \tag{44}$$

$$U(k, X; y) = k^a X. \qquad (a,b > 0) \tag{45}$$

(Later, k will be replaced by $u(k, \alpha)$ in eq (45).)

Eq (44) gives us a rectangular distribution of incomes as in Fig 15, and eq (45) is just a Cobb-Douglas utility indicator. From eq (45) we obtain the marginal rate of substitution Ω, as follows:

$$\Omega(k, R(k), y) \equiv \frac{U_k(k, y - R(k); y)}{U_X(k, y - R(k); y)}$$

$$= \frac{a(y - R(k))}{\bar{k}} \tag{46}$$

With these simplifications we get from eq (43):

$$y(k) = \underline{y} + b\left(1 - \frac{1}{T}(t^* + \frac{1}{\delta}\log(\bar{k}/k))\right) \quad \text{for } all\ \underline{k} \leq k \leq \bar{k},$$

and eq (6) becomes:

$$\frac{k}{a}R'(k) = \underline{y} + b\left(1 - \frac{1}{T}(t^* + \frac{1}{\delta}\log(\bar{k}/k))\right) - R(k)$$

This is a linear differential equation for R, and it is solved as follows:

$$\frac{d}{dk}(k^a R(k)) = ak^{a-1}\left(\frac{k}{a}R'(k) + R(k)\right)$$

$$= ak^{a-1}\left(\underline{y} + b(1 - \frac{1}{T}(t^* + \frac{1}{\delta}\log(\bar{k}/k))\right) \quad \text{by the equation.}$$

Both sides of this last equation can now readily be integrated between $\underline{k}$ and k:

$$k^a R(k) - \underline{k}^a R(\underline{k}) = a\int_{\underline{k}}^{k} dk'\,(k')^{a-1}\left(\underline{y} + b(1 - \frac{1}{T}(t^* + \frac{1}{\delta}\log(\bar{k}/k)))\right) \tag{47}$$

Here we must notice that an indefinite integral of $k^{a-1}\log k$ is the function $(1/a^2)k^a(a\log k - 1)$. (This can be checked by differentiation.) Making use of this result, and recalling eq (42), we obtain:

$$k^a R(k) - \underline{k}^a R(\underline{k}) = (k^a - \underline{k}^a)(y - \frac{b}{\delta T}(\log \underline{k} + \frac{1}{a}))$$

$$+ \frac{b}{\delta T}(k^a \log k - \underline{k}^a \log \underline{k}),$$

whence:

$$R(k) = (\underline{k}/k)^a R(\underline{k}) + (\underline{y} - \frac{b}{a\delta T})(1 - (\underline{k}/k)^a) + \frac{b}{\delta T} \log (k/\underline{k})$$

This, then, is the first of our unknown functions (for long-run equilibrium) and we may now directly check its properties. Differentiation gives:

$$R'(k) = \frac{1}{k}\{a(\underline{k}/k)^a(\underline{y} - R(\underline{k}) - \frac{b}{a\delta T}) + \frac{b}{\delta T}\}$$

$$= \frac{1}{k}\{a(\underline{k}/k)^a(\underline{y} - R(\underline{k})) + \frac{b}{\delta T}(1 - (\underline{k}/k)^a)\} \tag{48}$$

It is clear at once that $R'(k)$ is always positive and monotonically decreasing, as one would wish. ($\underline{y} - R(\underline{k})$ is positive because it is the X chosen by the lowest-income person - the part of his income not spent on housing in fact.) Next we can examine condition (39). We have that

$$R'(\bar{k}) = \frac{e^{-\delta(T-t^*)}}{\underline{k}} \{\left(a(\underline{y} - R(\underline{k})) - \frac{b}{\delta T}\right) e^{-a\delta(T-t^*)} + \frac{b}{\delta T}\}$$

This is expressed in terms of t^*, and for condition (39), either t^* must be put in terms of c or *vice versa*. From eq (15) the link is found:

$$c = \int_0^{T-t^*} dt' \, e^{-(\rho+\delta)t'} R'(\underline{k}e^{\delta(T-t^*-t')}) \tag{49}$$

It will be in order then, to use t^* as the exogenous parameter that must satisfy conditions which will allow condition (39) to hold. By use of eq (38), eq (49) can be evaluated, and the result is:

$$(\rho + \delta)c = \frac{e^{-\delta(T-t^*)}}{\underline{k}} \{\left(a(\underline{y} - R(\underline{k})) - \frac{b}{\delta T}\right) e^{-a\delta(T-t^*)}$$

$$\times \frac{\rho + \delta}{a\delta - \rho} (e^{(a\delta-\rho)(T-t^*)} - 1) + \frac{b}{\delta T}(1 + \frac{\delta}{\rho})(1 - e^{-\rho(T-t^*)})\}$$

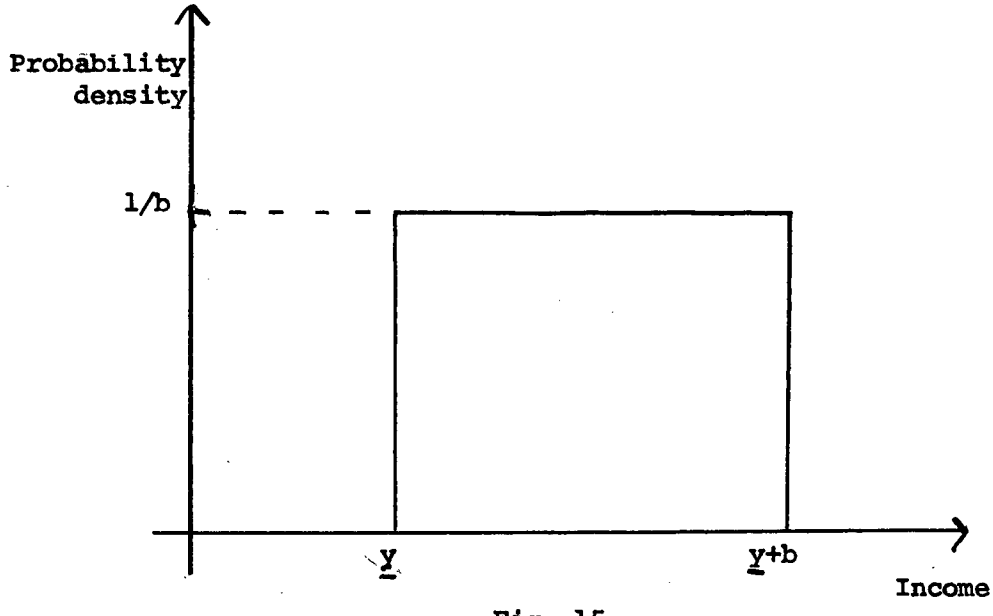

Probability density
1/b
y
y+b
Income

Fig 15.

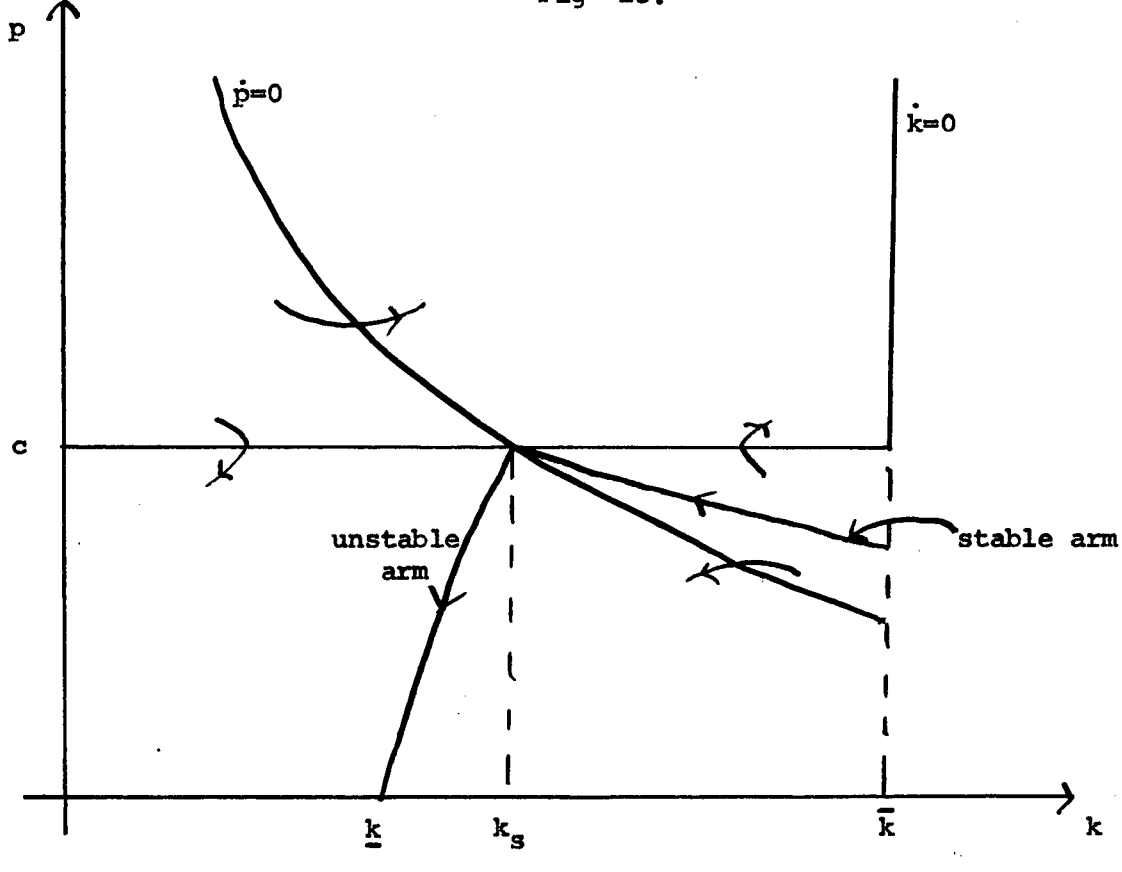

p
ṗ=0
k̇=0
c
unstable arm
stable arm
k
k_s
k̄
k

Fig 16.

Plainly a sufficient condition for inequality (39) to hold is that

$$\frac{\rho + \delta}{a\delta - \rho}\left(e^{(a\delta-\rho)(T-t^*)} - 1\right) < 1$$

and $$(1 + \frac{\delta}{\rho})(1 - e^{-\rho(T-t^*)}) < 1$$

Sufficiently small $T - t^*$ (which is equivalent to sufficiently small c) allows both of these inequalities to be satisfied. A little manipulation shows that a condition which approximates both inequalities is

$$(\rho + \delta)(T - t^*) < 1$$

It is of some interest to investigate the consequences of a choice of exogenous quantities such that condition (39) is *not* satisfied. If the marginal revenue function R' is used as calculated in eq (48) and the phase plane drawn, the result will look like Fig 16. But there is of course no reason for this not to be quite correct. Furthermore, none of the above analysis needs to be changed, except to replace $\bar{k}$ by k_s, the saddlepoint value of k. This follows because the initial point of the optimal upkeep path must still lie on the $\dot{k} = 0$ line, and the only feasible path ending at $\underline{k}$ is one which consists of staying at the saddlepoint until age t* (possible since I(c) is indeterminate and may be set equal to δk_s.) and then exiting along the unstable arm, reaching the k-axis at age T. Another point emerges clearly from this discussion. It was stated in Chapter IIIA that treating $\underline{k}$ as exogenous was tantamount to writing down an exogenous construction-site marginal cost schedule. But here, once c and $\bar{k}$ (the point at which the schedule goes to infinity) are given, our assumption has been that the schedule was known. Consequently, $\underline{k}$ should no longer be a parameter free to be chosen, and we can see in fact how it is to be determined.

Fig 16 shows that $\underline{k}$ must be at the end of the unstable arm, and it is this fact that pins $\underline{k}$ down. Eq (49) gives, for exogenous c, the quantity t* as a function of $\underline{k}$, since the function $R'(k)$, given by eq (48), involves only $\underline{k}$ and other exogenous parameters. ($R(\underline{k})$ is exogenous.) But for $\underline{k}$ to lie on the unstable arm, that is, the path leading out from the saddlepoint, we require that $\underline{k}e^{\delta(T-t^*)}$, which is the value of the variable k at time t* on the path ending at $\underline{k}$ at time T, should equal k_s, the saddlepoint k given by the equation $c(\rho + \delta) = R'(k_s)$. We require in fact that

$$c(\rho + \delta) = R'(\underline{k}e^{\delta(T-t^*(\underline{k}))}). \tag{50}$$

This is an equation for $\underline{k}$ in terms only of exogenous parameters.

We may now classify the cases that can arise by the use of eq (50). Let the solution of eq (50) be written as $\underline{k} = \underline{k}(c)$, and let $k_s = k_s(c)$ be just $\underline{k}e^{\delta(T-t^*(\underline{k}(c)))}$. Then if $k_s(c) \geq \bar{k}$, we get the case initially proposed, shown schematically in Fig 17(a) . The intermediate case $k_s(c) = \bar{k}$ is shown in Fig 17(b) , and the case $k_s(c) \leq \bar{k}$ is shown in Fig 16 . One case remains, and it is distinguished from the others not so much by the value of c as that of T. It could in principle happen (we shall not be much interested in this possibility) that T is so short that calculation gives a negative value for $t^*(\underline{k}(c))$. In this case, the optimal path starts at p = c as usual, but at a lower value of k than k_s - as shown in Fig 17(c) . We notice that in all cases no dwellings exist with a k greater than $\max(k_s, \bar{k})$, either because of technological impossibility or because of insufficient demand.

After this demonstration that our initial analysis will work in all cases with $\bar{k}$ redefined as $\max(k_s, \bar{k})$, we shall nonetheless stick with our

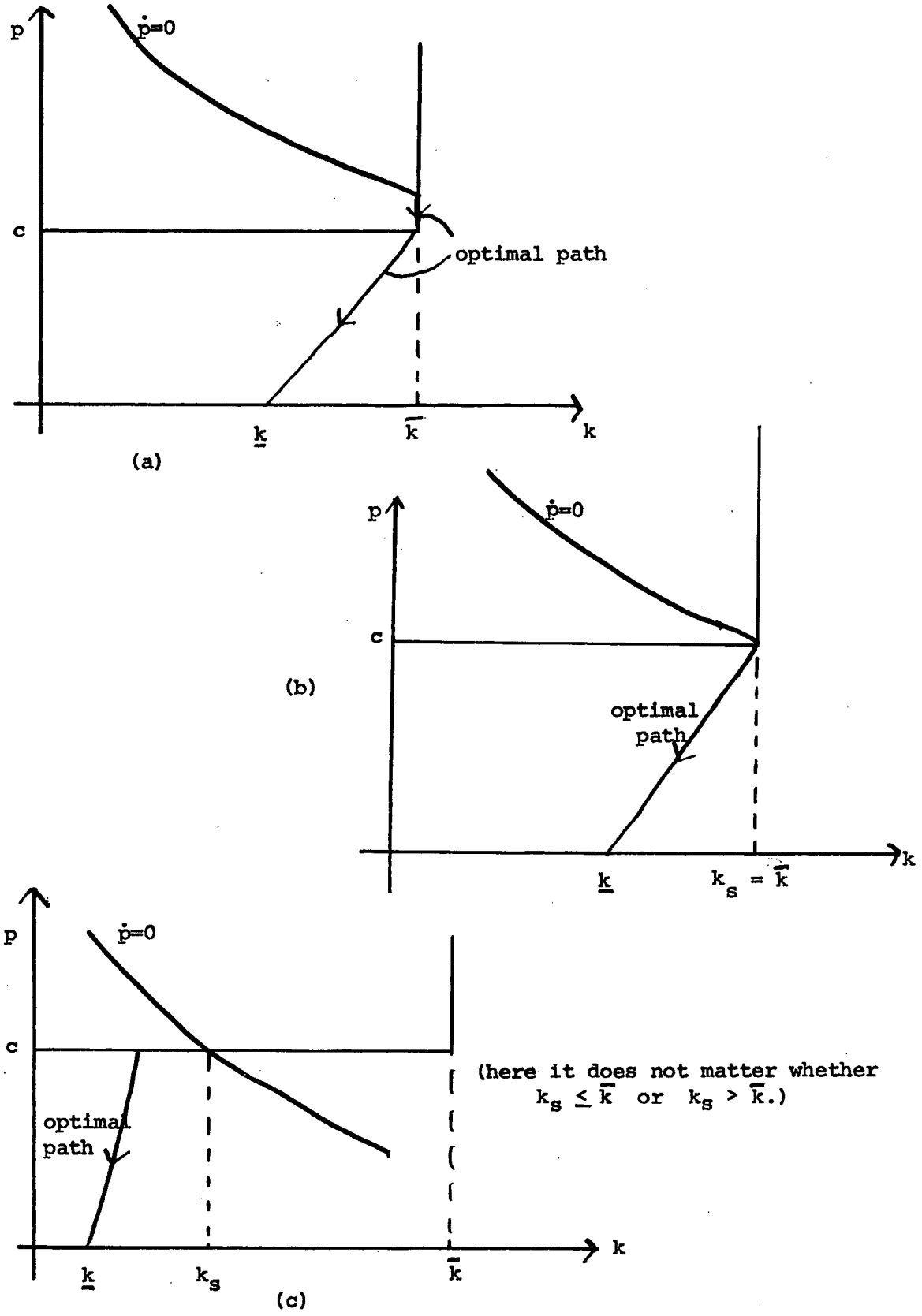

p
ṗ=0
c
optimal path
k
k̄
k
(a)
p
ṗ=0
c
(b)
optimal
path
k
k
$k_s = \bar{k}$
p
ṗ=0
c
(here it does not matter whether
$k_s \leq \bar{k}$ or $k_s > \bar{k}$.)
optimal
path
k
k
k_s
k̄
(c)

Fig 17.

interpretation of $\bar{k}$ as technologically imposed - for reasons of dynamics. Once the steady state is left, and demand becomes time-dependent, $k_s(c)$ will also be variable. It is much easier to deal with a technologically fixed $\bar{k}$. The last matter to be attended to in this section is to see that direct use of eq (26) gives the same result as the one we have obtained. This is now quite trivial. The operator T that appears in eq (26) is now equal simply to $\delta\bar{k}$ for $k = \bar{k}$, and zero for $k < \bar{k}$. The (unique) solution such that $t(\underline{k}) = T$ is immediate:

$$t(k) = t^* + \int_k^{\bar{k}} dk' \frac{1}{\delta k'} = t^* + \frac{1}{\delta} \log (\bar{k}/k) \quad \text{for } k < \bar{k}.$$

This is in accord with eq (40). For $k = \bar{k}$ the right-hand side of eq (26) is not defined, as we require to make sense of the result that $k(t) = \bar{k}$ for all t such that $0 \leq t \leq t^*$.

3. The Case of Rapid Suburban Construction.

This section presents the first part of the discussion of the dynamics of the coupled model of city and suburb. The suburb is modelled as in section 1 and the city as in section 2. Specifications of some more exogenous functions are made, and the meaning of the title of this section, "rapid suburban construction", is given. Then, the rents function R(k, t), now a function of time as well as of comfort, is obtained in terms of the supply function G(k, t) *via* the demand differential equation. The supply function G(k, t) is next determined by consideration of the city landlords' optimal control problem. A particular case of the dynamical evolution, that in which all urban maintenance stops immediately after the externality appears, is treated first. The function G is then easily written down, and consequently also the rents

function, R. A series of checks has then to be undertaken to determine when the particular case applies - these checks are concerned with the dynamics both of the city and of the city-suburban margin. Other possible régimes of dynamical evolution are discussed following the checks. Lastly, it is verified that the evolution of the model is *stable*, in the sense that it is indeed the richer city-dwellers who first become dissatisfied and move out to the suburbs.

First then, let us specify in a particularly simple form the function f which appears in eq (30) and which links the rate of construction in the suburbs, m(t), to the profit per dwelling, $\pi(t)$. Let

$$f(\pi) = \begin{cases} m & \text{if } \pi > 0 \\ 0 & \text{if } \pi < 0 \end{cases} \tag{51}$$

and f(0) is indeterminate between zero and m. This means (eq (32)) that the number of dwellings in the suburbs at time t is M(t) = mt for so long as positive profits exist over an unbroken time interval. What eq (51) says is just that there is a perfectly inelastic response of entry by exactly m firms for any positive profit whatever, and instant exit of all of them in the face of loss. At the margin of exactly zero profit, there may be any number between zero and m. This choice of the function f makes our calculations much simpler than would any other choice, and does not obscure the dynamical questions in which we are principally interested.

The rate at which demolition goes on in the city is, at least for time T after the imposition of the externality, given by N/T. (Long-run equilibrium, with a rectangular distribution of building ages, prevails before this time.) For this section, we consider only the case m > N/T, that is, the case in which positive profits in suburban construction cause more dwellings to be built there than are simultaneously being demolished in the city. The result is, of course, more dwellings than people, and so

there are unoccupied dwellings in the city. We may now specify the form of the function $u(k, \alpha)$, which expresses how a comfort level enters a tenant's utility function in the presence of the externality α. A particularly easy form is

$$u(k, \alpha) = \alpha k. \tag{52}$$

For this choice of u, then, and with the marginal rate of substitution given by eq (46), the demand equation in the city, eq (33), becomes:

$$R'(k, t) = \frac{a(y - R(k, t))}{k} \tag{53}$$

just as before. The demand equation is *not* unchanged in general - that is just a felicitous result of eq (52). Time dependence has been made explicit, and dashes denote differentiation with respect to the upkeep variable.

The fact of unoccupied dwellings means that the rent charged on the least comfortable inhabited dwelling is zero - dwellings any less comfortable have become free goods. This observation provides the boundary condition to accompany eq (53). The zero-profit condition no longer applies of course, since we assume that the externality arrives unexpectedly. Let the lowest upkeep level of any inhabited dwelling at time t be $k_{out}(t)$. Then $R(k_{out}(t), t) = 0$ and so eq (53) gives:

$$R(k, t) = ak^{-a} \int_{k_{out}(t)}^{k} dk' \, (k')^{a-1} \, y(k', t)$$

where $y(k, t)$ is, as usual, the income of the person inhabiting a dwelling of upkeep level k at time t.

Next, let us define the upkeep supply function $G(k, t)$ as follows: $NG(k, t) \equiv$ number of dwellings in existence (not necessarily inhabited) at time t of upkeep level $\leq k$. We may express $k_{out}(t)$ in terms of G.

Since the number of uninhabited dwellings at time t is just $(m - (N/T))$, we have

$$NG(k_{out}(t), t) = (m - (N/T))t. \tag{54}$$

Similarly, equating numbers of dwellings to numbers of people, one obtains for $y(k, t)$ the equation:

$$NF(y(k, t)) + (m - (N/T))t = NG(k, t),$$

so that

$$y(k, t) = F^{-1}\{G(k, t) - ((m/N) - (1/T))t\},$$

whence

$$R(k, t) = ak^{-a} \int_{k_{out}(t)}^{k} dk' \, (k')^{a-1} \{\underline{y} + b\left(G(k', t) - (\frac{m}{N} - \frac{1}{T})t\right)\} \tag{55}$$

by use of eq (44).

We have now reached the stage where, if we can find $G(k, t)$, the problem is done. To find $G(k, t)$, we must, as always, consider our optimal control problem for landlords. First, since at time zero (the moment of imposition of the externality) $G(k, 0)$ is just the long-run equilibrium function given by eq (41), we may see from eq (55) that $R(k, 0)$ is less than the equilibrium $R(k)$ by just $\underline{k}^a R(\underline{k})/k^a$, a quantity which is always positive. (see eq (47).) On the other hand, a decrease in $R(k, 0)$ means, because of eq (53), an *increase* in $R'(k, 0)$, so that the impact effect of the externality is "perverse", in the sense of Chapter II, in that the total revenue schedule falls, but the marginal revenue schedule rises. There is nothing perverse economically of course: the lower rents leave more money for other things, and the marginal rate of substitution shifts in favour of housing. The impact effect is not, of course, the whole story. If it were, then landlords would tend to maintain dwellings to a *greater* age than in equilibrium - eq (49) shows that larger R' means a shorter time interval $T - t^*$.

There are in fact circumstances in which the landlords optimal

response is to cease all maintenance, and we shall now consider this case. Any dwelling of age v less than T - t* (t* will throughout denote the equilibrium t* given by eq (49)) is at time t = 0 in state $\bar{k}$. It will finish its life at time T - v, at which time the shadow price of upkeep, p(t), will be zero. At time t = 0, then, we get from eq (10) that

$$p(0) = \int_0^{T-v} dt\ e^{-(\rho+\delta)t}\ R'(k(t), t)$$

where k(t) is the upkeep at time t. If no maintenance is done for t > 0, then $k(t) = \bar{k}e^{-\delta t}$, and so

$$p(0) = \int_0^{T-v} dt\ e^{-(\rho+\delta)t}\ R'(\bar{k}e^{-\delta t}, t) \tag{56}$$

But *no* dwelling can reach age T and still be receiving a positive rent. Further, there is always a *positive* time interval during which no rent is received and for which therefore R′ = 0, since R(k, t) = 0 for all $k < k_{out}(t)$. The upper limit on the integral in eq (56) can thus be extended to ∞, and it is clear at once that p(0) is the *same* for all dwellings of age less than T - t*. If, then, p(0) as given by eq (56) is less than c, there will indeed be no maintenance after t* = 0.

Let us now complete the analysis of this case. It is immediate that

$$NG(k, t) = NG(ke^{\delta t}, 0) - Nt/T \quad \text{for } k < \bar{k}e^{-\delta t}$$

and $NG(\bar{k}e^{-\delta t}, 0) = NG(\bar{k}, 0) - Nt/T.$

From eq (41) we obtain:

$$G(k, t) = 1 - \frac{1}{T}(t^* + \frac{1}{\delta}\log(\bar{k}/ke^{\delta t})) - \frac{t}{T}$$

$$= 1 - \frac{1}{T}(t^* + \frac{1}{\delta}\log(\bar{k}/k)) \text{ for } k < \bar{k}e^{-\delta t}, \text{ just as before} \tag{57}$$

and $G(\bar{k}e^{-\delta t}, 0) = 1 - t/T.$

The function $k_{out}(t)$ comes from eq (54) and is:

$$k_{out}(t) = \bar{k}e^{-\delta\{T-t^*-t((mT/N)-1)\}}. \tag{58}$$

Of course this makes sense only if $k_{out}(t) < \bar{k}e^{-\delta t}$, since the right-hand side here is, at time t, the greatest existing upkeep level. This means that $T - t^* - mTt/N > 0$, i.e. that $t < \frac{N}{mT}(T - t^*)$. Once $t = \frac{N}{mT}(t - t^*)$, the free-good upkeep level coincides with the greatest upkeep level, and all city housing is free. A simple, but tedious, calculation of eq (55) gives R(k, t). The result is:

$$R(k, t) = \left(\underline{y} - \frac{b}{a\delta T}\right)\left(1 - (k_{out}(t)/k)^a\right) + \frac{b}{\delta T}\log(k/k_{out}(t)) \tag{59}$$

From this one can calculate eq (56). The upper limit of the integral is the time $\bar{t}$ for which $\bar{k}e^{-\delta\bar{t}} = k_{out}(\bar{t})$, which means that

$$\bar{t} = (N/mT)(T - t^*). \tag{60}$$

The answer is:

$$p(0) = (a/\bar{k})\left(\underline{y} - \frac{b}{a\delta T}\right)e^{-a\delta(T-t^*)}\frac{1}{(a\delta mT/N) - \rho}$$

$$\times\left(e^{((a\delta mT/N)-\rho)N(T-t^*)/mT} - 1\right)$$

$$+ \frac{b}{\rho\delta T\bar{k}}\left(1 - e^{-\rho N(T-t^*)/mT}\right) \tag{61}$$

Thus, finally, if this quantity is less than c, there will indeed be no maintenance after $t = 0$, and eqs (57), (58) and (59) provide the solution to the problem. After time $t = N(T - t^*)/mT$, no more rent can be collected in the city, and property continues to deteriorate exponentially.

For this fearsome tale to proceed to the end, that is, complete desertion of the city, it is necessary that profits remain non-negative in the suburbs. The continued lack of maintenance in the city means that

the maximum utility to be had there steadily declines. Therefore, for as long as anyone left in the city can afford it, there will be a steadily growing incentive to move to the suburbs. Even if, after some time, profits there fall to zero, a rate of construction less than m can be maintained.

Let us now choose a specific form for the suburban construction cost function $\Xi(\kappa, M)$:

$$\Xi(\kappa, M) = C(1 + hM) + \ell\kappa. \qquad (h, \ell, C > 0)$$

This comprises a fixed building cost $C(1 + hM)$ which grows with M, the number of already existing suburban buildings, and a linear variable cost $\ell\kappa$, dependent on the built-in standard of comfort. Eq (36) can be solved with this choice of the function Ξ and the other choices made previously. Eq (36) becomes:

$$\ell = \frac{1}{\kappa}\left(a\alpha_s(y - \pi - C(1 + hM) - \ell\kappa)\right)$$

with solution $$\kappa = \frac{a}{\ell(a + 1)}(y - \pi - C(1 + hM)).$$

From this we may calculate the utility obtainable in the suburbs by a person of income y, when the profit there is π and there are M dwellings in existence. We obtain (eq (45)):

$$U_{sub} = \frac{a\alpha_s}{\ell(a + 1)^2}(y - \pi - C(1 + hM))^2 \tag{62}$$

where it is necessary that $y > \pi + C(1 + hM)$ in order that the rent paid does not exceed total income. For a city-dweller of income y who at time t avails himself of the highest obtainable standard of comfort, utility attained is:

$$U_{city} = \alpha\bar{k}e^{-\delta t}(y - R(\bar{k}e^{-\delta t}, t)). \tag{63}$$

From eqs (58) and (59) one has:

$$R(\bar{k}e^{-\delta t}, t) = \left(\underline{y} - \frac{b}{a\delta T}\right)\left(1 - e^{-a\delta(T-t^*-(mtT/N))}\right) + \frac{b}{T}\left(T - t^* - \frac{mtT}{N}\right) \tag{64}$$

Now, while the profit π is positive, we know that $M = mt$. The income of the person at the margin of indifference between city and suburban living is then (eq (37) and eq (44)):

$$F^{-1}(1 - (mt/N)) = \underline{y} + b(1 - (mt/N)).$$

Therefore the marginal condition, eq (38), which gives π, can be written down by equating the right-hand sides of eqs (62) and (63). The result is, at time t:

$$\frac{a\alpha_s}{\ell(a+1)^2}\{\underline{y} + b(1 - (mt/N)) - \pi - C(1 + hmt)\}^2 \tag{65}$$

$$= \alpha\bar{k}e^{-\delta t}\left\{\left(\underline{y} - \frac{b}{a\delta T}\right) e^{-a\delta(T-t^*-(mtT/N))} + \frac{b}{T}\left(t^* + \frac{1}{a\delta}\right)\right\}$$

whence

$$\pi(t) = \underline{y} + b(1 - (mt/N)) - C(1 + mht) - (a+1)\left(\frac{\ell\alpha\bar{k}e^{-\delta t}}{a\alpha_s}\left\{\left(\underline{y} - \frac{b}{a\delta T}\right) e^{-a\delta(T-t^*-(mtT/N))} + \frac{b}{T}\left(t^* + \frac{1}{a\delta}\right)\right\}\right)^{\frac{1}{2}} \tag{66}$$

This is clearly a decreasing function of t. In those cases, then, where the city does not simply empty directly, a time, t_0, say, will be reached when π becomes zero. This time t_0 is calculated by setting the right-hand side of eq (66) equal to zero.

After this, it is the fact that $\pi = 0$ if any more suburban construction occurs that determines the dynamics. Let us, for the sake of simplicity,

assume that $t_0 > \bar{t}$ ($\bar{t}$ is the time when rents in the city go to zero - eq (60)). If not, the following analysis will be more complicated, but still feasible. With $t_0 > \bar{t}$, the marginal condition, eq (38), when $\pi = 0$, reads:

$$\frac{a\alpha_s}{\ell(a+1)^2}\{\underline{y} + b(1 - M/N) - C(1 + hM)\}^2$$

$$= \alpha\bar{k}e^{-\delta t}\{\underline{y} + b(1 - M/N)\} \tag{67}$$

and this can be solved directly for M as a function of t. We must notice that since the right hand side of this equation is always positive, M can never exceed the value, $\bar{M}$, say, which makes the left-hand side zero:

$$\bar{M} = \frac{N(\underline{y} + b - C)}{b + hNC} .$$

If $\bar{M} < N$, there is a fraction of the city population which can never afford suburban housing, and is left trapped in the decaying city, albeit with free housing. Eventually, demolition will catch up with these people, and housing will no longer be free. Our present model does not say what will happen then, but it must certainly be something breaking the pattern of preceding events. (Government subsidies, changed municipal rules, riots, illegal squatting are all possibilities.)

There is one more matter to be checked before this section is concluded. It has been assumed all along that the margin between city and suburban living was such that people with incomes greater than the marginal one preferred the suburb, people with lower incomes the city. This assumption must be verified. At time t, a person with income ε less than the marginal one will attain a utility less than that of the marginal person by $\Delta U_{city} = \alpha\bar{k}e^{-\delta t}\varepsilon$ (compare eq (63)).

or by $\Delta U_{sub} = \dfrac{2a\alpha_s\varepsilon}{\ell(a+1)^2}(y - \pi - C(1 + hM))$. (compare eq (62))

We require then that

$$\frac{2a\alpha s}{\ell(a+1)^2}(y - \pi - C(1 + hM)) > \alpha\bar{k}e^{-\delta t}.$$

If in this inequality, $\alpha\bar{k}e^{-\delta t}$ is replaced by the expression for it obtained from the marginal condition itself, eq (65), the requirement becomes:

$$2(y - R(\bar{k}e^{-\delta t}, t)) > y - \pi - C(1 + hM),$$

i.e. $y + \pi + C(1 + hM) > 2R(\bar{k}e^{-\delta t}, t)$.

For times $t > \bar{t}$, this is trivially satisfied. Inspection of eq (64) shows that $2R(\bar{k}e^{-\delta t}, t)$ decreases faster with t than does ChM, and so the condition will certainly be satisfied for all t if it is at $t = 0$. A sufficient condition is then

$$\bar{y} + C > 2R(\bar{k}, 0),$$

which will always be satisfied in normal circumstances. The richest person would otherwise be spending well over half his income on rent *after* the externality - a most unlikely state of affairs. We may conclude then that the course of events is as described.

4. The Case of Slow Suburban Construction.

In the last section, we considered some of the possible outcomes in the event that there were unoccupied dwellings in the city. The state of affairs is quite different if there are not. The reason for this is simple: if one excludes the knife-edge case in which the rate of suburban construction exactly equals the rate of urban demolition, the former must be less than the latter if no unoccupied dwellings exist, and so, if everyone is to be housed somewhere, urban construction must still be going on and must be profitable. We may assume that the profit for urban construc-

ion is exactly zero - this implies competition and no barriers to entry.

First, the effects of continued urban construction can be taken into account so as to provide the distribution of dwelling *ages* (as opposed to that of *upkeep* levels) at any time. The distribution of upkeep levels can then be deduced by introducing a function $s(t)$, which gives the time at which maintenance on a dwelling constructed at time t comes to an end. As usual the rents function $R(k, t)$ follows from the demand differential equation. The next step in the analysis is to obtain an equation for $s(t)$ from the landlords' optimal control problem. This equation is unfortunately rather involved, and it is not possible to provide an explicit solution. However an iterative scheme is described by which it may be computed. Lastly, it is pointed out that a much simpler result can be obtained if the strict rational expectations hypothesis is relaxed.

We assume in this section, that $m < N/T$. Now the urban demand equation, (53), is the same as before, but its boundary condition is no longer that $R(k_{out}(t), t) = 0$, but rather the zero-profit condition. It should be recalled that we have assumed that urban landlords have perfect foresight, and that their profit is then to be calculated with the rent function $R(k, t)$ predicted by the model. At time $t = 0$ we shall have suburban profit $\pi > 0$, if any move to the suburbs is to take place at all. Presumably after some time π will fall to zero, and then, since urban construction is continuing and housing of upkeep level $\bar{k}$ is still available, suburban construction will permanently cease. The only thing that would make it start up again would (other than another exogenous adverse externality) be an increase in city rents. Since there are never unoccupied dwellings, the zero-profit condition applies to every building put up after $t = 0$, and thus $R(\bar{k}, t)$ will be bounded above and

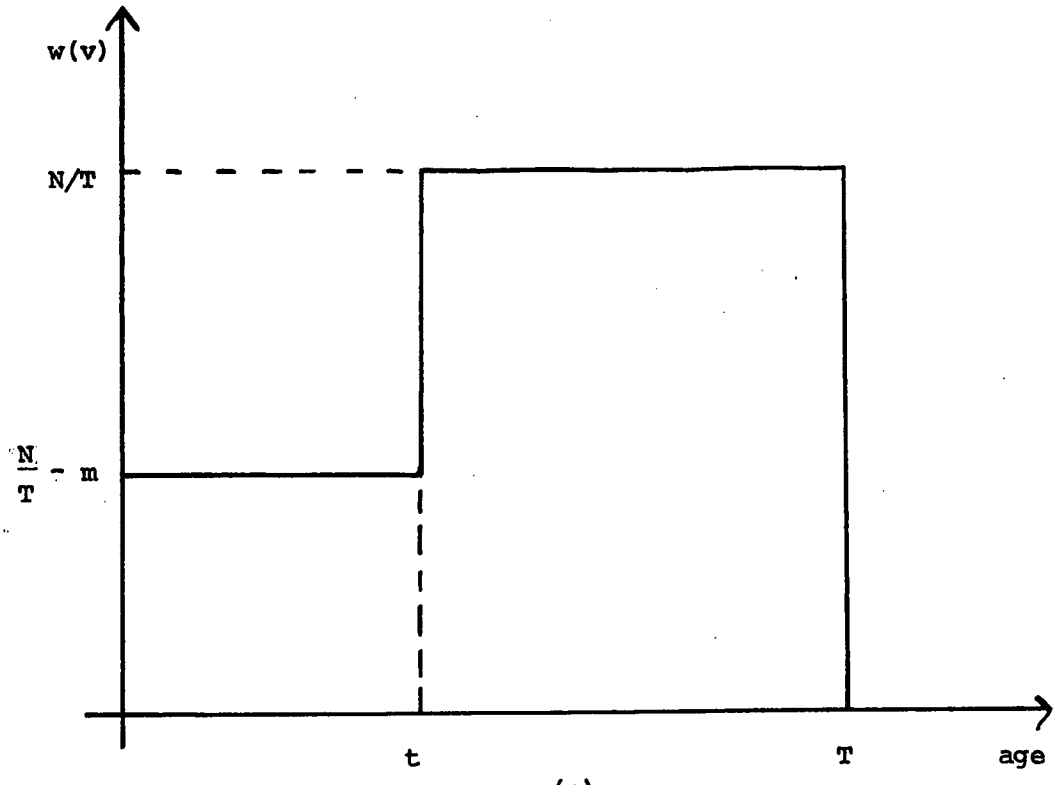
below - only buildings in existence at $t = 0$ can incur windfall gains or losses. With our assumption that people who move to the suburbs never move back to the city, the state of affairs where the richest person left in the city is just indifferent to moving to the suburbs when $R(\bar{k}, t)$ is at its highest point is stable: no more new suburban construction will ever take place (except for replacement of course - we have abstracted from such considerations).

It may seem that there is no reason for $R(\bar{k}, t)$ to change at all from its $t < 0$ value. Indeed there is no reason for it to change by very much, but there will be small fluctuations, as we shall now see. While $\pi > 0$, the rate of urban construction falls from N/T to $(N/T) - m$. This means that the distribution of building ages is no longer rectangular. In fact, for time t, the distribution will be as in Fig 18(a). If, as we may for simplicity assume, once π reaches zero suburban construction stops for good, urban construction rises again to a rate N/T, and afterwards the age distribution will be as in Fig 18(b). Let us denote the function graphed in Fig 18 by $w(v, t)$, and then $w(v, t)\, dv$ is the number of dwellings which at time t have an age between v and $v + dv$. After suburban construction ceases, at time $t^{\dagger}$ say, we have

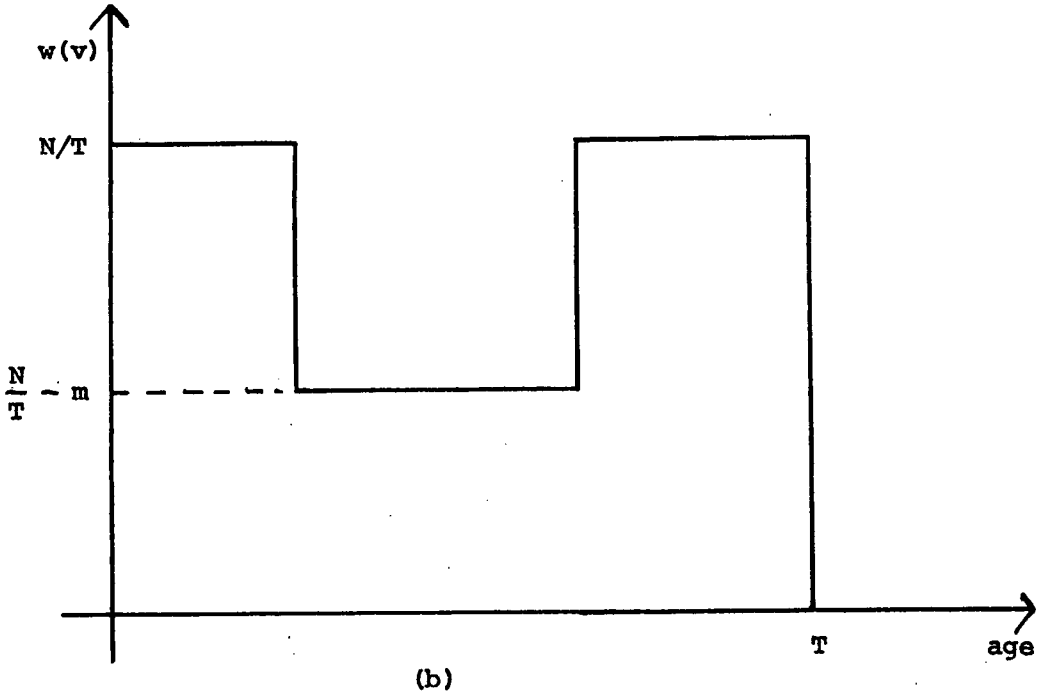
$$w(v, t) = w(v + t^{\dagger} - t + nT, t^{\dagger}) \tag{68}$$

where n is an integer chosen so that $0 \leq v + t^{\dagger} - t + nT \quad T$.
It is as though the function $w(v, t^{\dagger})$ were reproduced in each interval of length T as a periodic function, and then propagated itself forwards like a wave. If there ar finally $\bar{M}$ buildings in the suburbs, construction proceeds so as to keep $N - \bar{M}$ buildings in the city: construction rate after $t = t^{\dagger}$ always equals demolition rate.

Whether the function $w(v, t)$ takes on only the values N/T and



(a)



(b)

Fig 18.

(N/T) - m as shown in Fig 18 or has some intermediate values attained while $R(\bar{k}, t)$ rises to its highest point, let us now make the definition:

$V(v, t) \equiv$ number of buildings with age $\geq v$ at time t.

The function V is of course calculated directly from the function w, but it is more convenient to work with V in the analysis to follow.

It can now be seen why the maximum city rent, $R(\bar{k}, t)$, fluctuates. There will be periods when older housing is scarcer than at others. Consequently the upkeep supply function, $G(k, t)$, will at times increase more slowly with k than at others. This means that the whole rents profile, $R(k, t)$, will assume different shapes at different times, and since the value of a building put up at time t depends on rents from time t to time t + T in such a way that, with unchanging construction costs, exactly zero profit is realised, it follows that $R(\bar{k}, t)$ cannot always have the same value. It is reasonable to suppose that its fluctuations will be minor - they are certainly bounded - and, clearly, if a long-run equilibrium with a rectangular distribution is ever to be achieved, frictional forces of some kind must operate so as to damp out both the fluctuations of $R(\bar{k}, t)$ and those of $V(v, t)$.

We may now finish the present analysis. Let the time at which a building constructed at time t ceases to be maintained at upkeep level $\bar{k}$ be denoted by s(t). This time is then determined by the time-dependent analogue of eq (15):

$$c = \int_{s(t)}^{t+T} dt' \, e^{-(\rho+\delta)(t'-s(t))} \, R'(\bar{k}e^{-\delta(t'-s(t))}, t') \tag{69}$$

(c is the marginal cost of upkeep - this equation the the condition that $p(s(t)) = c$, $p(t + T) = 0$, where p is the shadow price of upkeep.) The differential equation for the rent function $R(k, t)$ is eq (53), and its

solution may be written as follows:

$$R(k, t) = (\bar{k}/k)^a R(\bar{k}, t) - \frac{a}{k^a} \int_k^{\bar{k}} dk' \, (k')^{a-1} \, y(k', t) \tag{70}$$

where $y(k, t)$ is as usual the income of the person choosing comfort level k at time t. The function $y(k, t)$ can be expressed in terms of $V(v, t)$ and $s(t)$. A building constructed at time t', is, at time $t > s(t')$, in a state of upkeep $\bar{k}e^{-\delta(t-s(t'))}$. The number of buildings in a less good state of upkeep is $V(t - t', t)$ - that is, the number of buildings of age greater than $t - t'$, at time t. Equating numbers of people and dwellings as usual, we obtain:

$$y(\bar{k}e^{-\delta(t-s(t'))}, t) = F^{-1}\left(\frac{1}{N} V(t - t\ , t)\right)$$

$$= \underline{y} + \frac{b}{N} V(t - t', t)$$

by eq (44). Therefore:

$$y(k, t) = \underline{y} + \frac{b}{N} V\left(t - s^{-1}(t - \frac{1}{\delta} \log (\bar{k}/k)), t\right) \tag{71}$$

where s^{-1} is the function inverse to s. When this is substituted into eq (70), the function $R(k, t)$ is expressed in terms of $s(t)$ and $V(v, t)$. But $V(v, t)$ can be regarded as known, since it can be calculated directly from the marginal condition of indifference between city and suburb. Finally then, to complete the set of dynamical equations, there is the zero-profit condition. If B denotes the unchanging cost of a new building in state $\bar{k}$, then this condition is:

$$e^{-\rho t} B = \int_t^{s(t)} dt \quad e^{-\rho t'} (R(\bar{k}, t') - c\delta\bar{k}) \quad + J(t), \tag{72}$$

where $J(t)$ is the value to be obtained from the building once maintenance

stops, that is:

$$J(t) = \int_{s(t)}^{t+T} dt' \, e^{-\rho t'} R(\bar{k}e^{-\delta(t'-s(t))}, t') \quad . \tag{73}$$

Since R(k, t) is already expressed in terms of s(t), so then is J(t) by eq (73) and so therefore is the whole of eq (72). Eq (72) is then an equation of a very involved kind for the function s(t).

It would be exceedingly difficult and probably pointless to try to obtain an explicit expression for s(t), even with the many simplifying assumptions so far made. But, as was pointed out before, the fluctuations in $R(\bar{k}, t)$ are small in any normal circumstances. An iterative scheme for solving eqs (72) and (69) may therefore be suggested, suitable of course only for numerical computation. A reasonable first guess for the function s(t) is $s(t) = t + (T - t^*)$,
where t^* is the long-run equilibrium parameter given by eq (49). The city-suburb marginal condition is just

$$\frac{a\alpha_s}{\ell(a+1)^2} \left(\underline{y} + b(1 - (mt/N)) - \pi - C(1 + hmt)\right)^2$$
$$= \alpha\bar{k} \left(\underline{y} + b(1 - (mt/N)) - R(\bar{k}, t)\right) \tag{74}$$

byanalogy with eq (65), and if one sets $R(\bar{k}, t)$ equal to $R(\bar{k})$, the pre-externality equilibrium value, and sets $\pi = 0$, an estimate is obtained for the time t at which suburban construction stops. From this the function V(v, t) is calculated using eq (68). Then the function y(k, t) follows from eq (71), and hence R(k, t) from eq (70). The quantity J(t) can then be obtained from eq (73) and then eq (72) is an integral equation for $R(\bar{k}, t)$. It can be solved by differentiating with respect to t:

$$-\rho e^{-\rho t} B = e^{-\rho s(t)} (R(\bar{k}, s(t)) - c\delta\bar{k}) - e^{-\rho t} (R(\bar{k}, t) - c\delta\bar{k}) + J'(t) .$$

This can be solved for $R(\bar{k}, t)$ in terms of $R(\bar{k}, s(t))$. One may then iterate and write $R(\bar{k}, s(t))$ in terms of $R(\bar{k}, s(s(t)))$ and so forth; the exponential factors ensure convergence. With the solution $R(\bar{k}, t)$ in hand, a new expression for $R(k, t)$ comes from eq (70). This can be substituted into eq (69) so as to obtain a second-round estimate of $s(t)$. The whole procedure may be started again, and will in all likelihood converge very rapidly.

The extreme complication of this result is due to our rational expectations hypothesis. If landlords are endowed with a little less foresight, then things become much simpler. Let us imagine, for instance, that landlords cannot be persuaded to invest in new construction unless the construction cost, B, is covered by the return

$$\int_t^{t+T} dt' \, e^{-\rho(t'-t)} (R(\bar{k}, t') - c\delta\bar{k})$$

which would be obtained if full maintenance continued to the time of demolition. This return of course must be less than that really obtainable, since it is feasible but not optimal. Then eq (72) becomes just

$$B = \int_t^{t+T} dt' \, e^{-\rho(t'-t)} (R(\bar{k}, t') - c\delta\bar{k}),$$

and this equation is satisfied by a constant value of $R(\bar{k}, t)$, $R(\bar{k}, t) = R$, say. The city-suburb marginal condition , eq (74), is then satisfied by a unique $t = t^{\dagger}$ at which $\pi = 0$, and no further suburban construction at all occurs for $t > t^{\dagger}$. The age distribution, $V(v, t)$, is then known completely. The function $s(t)$, which gives the time at which a building constructed at time t is *in fact* left to decay has still to be calculated

with some difficulty from eqs (69), (70) and (71), but of course it is no longer of much interest.

The behaviour of s(t) in a general way may be seen by inspection. When older buildings are scarce, that is, when buildings put up in times of urban construction rate (N/T) - m are nearing the end of their lives, then for any v, $V(\bar{v}, t)$ is smaller than usual. From eq (71) it follows that y(k, t) is smaller than usual for any k, and therefore by eq (70) that R(k, t) is greater. From eq (53), $R'(k, t)$ is smaller, and so from eq (69) s(t) is nearer to t and further from t + T than usual. In other words, the equations operate so as to mitigate the scarcity of older buildings by causing maintenance to come to an end sooner.

This concludes our formal analysis of the city-suburb dynamics. It remains in the next section to summarise the numerous conclusions of this section and the preceding one, and to gather a few loose ends.

5. Summary and Conclusions.

It was not the aim in sections 3 and 4 to provide an exhaustive catalogue of everything that might happen in the two cases m > N/T and m < N/T. Rather, certain sequences of events were shown to be possible, and some details analysed. There were two reasons for this: the different cases that may arise for various choices of the model parameters are exceedingly numerous, so that an exhaustive catalogue would also be exhausting; and some of these cases, while very similar to each other, can be distinguished by large differences in complication or ease of analysis.

It is probably worthwhile, then, to supply here a verbal, rather

than mathematical, discussion of the information that can be obtained from our exercise in comparative dynamics.

In the rapid suburban construction case, $m > N/T$, it is not necessary that all city maintenance come to an end at $t = 0$. There is a condition for this, namely that the quantity $p(0)$ given in eq (61) should be less than c. What happens if this condition is not met? It was pointed out that eq (61) gives $p(0)$ for *every* building in state $\bar{k}$ at $t = 0$ and so it remains true that there is only one optimal maintenance path for all of them. Some time must pass, then, before general decay starts, and continuity requires that some buildings which had at $t = 0$ begun their decay should again be maintained for a time.

The time $\bar{t}$, given by eq (60), at which city housing becomes free, is not affected by maintenance after $t = 0$. It is determined only by the difference in the suburban construction rate and the urban demolition rate: it is the time when all the dwellings of age more than t^* at $t = 0$ have become unoccupied. Consequently it makes little difference to events after $t = \bar{t}$ whether or not $p(0)$ as given by eq (61) is less than c.

If there were no limit to the rate of suburban construction, the number of people in the suburbs, $M(t)$, would be given simply by the marginal condition with zero profit, eq (67). There would be an instant departure of a positive fraction of the population - $M(0)$ is not zero in general. (If a formal solution of eq (67) yields a *negative* value of $M(0)$, this means that suburban housing is too dear, and the urban externality, α, not severe enough relative to the suburban one, α_s, for anyone to wish to move out.) In any event, $M(t)$ as given by eq (67) specifies the greatest possible number of people in the suburbs at time t, so that in general this number will be min $(mt, M(t))$. The time t_o is

defined by $mt_o = M(t_o)$, i.e. by $\pi(t_o) = 0$.

It may happen that $M(t)$ falls below mt very quickly, so quickly in fact that after a time $\tilde{t}$, say, there are no more unoccupied dwellings in the city. The time $\tilde{t}$ would be defined by $M(\tilde{t}) = N\tilde{t}/T$. This equation says that the number of suburban dwellings at time $\tilde{t}$ equals the number of city dwellings demolished between $t = 0$ and $t = \tilde{t}$. But in this case urban construction would have to start again. The dire consequences mentioned in section 3, riots and so forth, need not take place if $\tilde{t}$ is small enough. For once urban construction starts, we are in the situation described in section 4, and the analysis presented there will go through in some cases.

At this point, we may see that a solution to the system of equations (72) and (69) may not exist. This will be so if the incomes of the people left in the city are not large enough for the rents obtainable from new buildings over their lifetime to cover costs. An extreme instance of this would be a state of affairs in which the highest income left in the city once suburban construction has ceased, $\underline{y} + b\bar{M}/N$, is less than the rent $R(\bar{k})$ which would in a steady state be necessary for eq (72) to be satisfied. Now city construction will of course fail to be profitable in much wider circumstances than this. It is then a state of affairs where no new city construction can be profitable that leads to the "dire consequences" of section 3.

Even in the case $m < N/T$ it can turn out that eqs (72) and (67) have no solution. If suburban housing is cheap enough, and the externality α severe enough, the marginal condition given by eq (74) can, when π is set equal to zero, lead to a value of mt sufficiently great that too few people are left in the city for new construction there to be profitable.

Again there will be dire consequences.

The model presented in this essay is not of course intended to provide a contribution to realistic urban economics. For this reason, I shall not attempt to make policy prescriptions for city managers or governments faced with population loss due to migration to suburbs, although, were the model realistic, many such prescriptions would be implicit in the analysis I have presented. On the other hand, I do hope that economic theory is in some small measure advanced by this essay. It has been shown how a general equilibrium model can be solved in an intertemporal context; the problems of entry into and exit from business have been explicitly incorporated into the analysis; some indication has been given of the richness of detail that comparative dynamics can provide when finite rather than infinitesimal changes in exogenous quantities are considered.

In the postscript that follows this collection of essays, there will be discussion of some of the many outstanding problems in comparative dynamics, and what one might do to try to solve them.

POSTSCRIPT.

Seeing into the future is an occupation fraught with hazards. It is ironic that a world more laden with statistics and projections, facts and figures than ever it was in the past is the one in which economic theorists are at last trying to come to grips with the effects of uncertainty and lack of foresight. It is a world, though, where much scientific research has still not told us enough about the life-cycles of fish or even of trees for poor economists to be able to take into account the facts of physical and biological evolution that are needed for anything that could be called *optimal* exploitation of renewable resources. It is a world where the estimates of what remains for us to use of non-renewable resources like oil and coal are as volatile as the stock market quotations. . .

It would be wrong of me to insist on how great a defect it is that this thesis does not worry about uncertainty. It *is* a defect, but one which can be cured with a bit of effort. Since von Neumann and Morgenstern (1944) and more especially Arrow's (1971) essays on risk-bearing, economists have made some progress in the matter. It may even be fair to say that the business of making any given economic model into a stochastic one is just technical. At all events, one knows how to tackle the problem. There are some other perhaps less obvious defects in the theory used in the thesis, and it is time to discuss them.

One such defect is that optimal control theory allows us to maximise only those objective functionals that are integrals. This is not really a

matter of too great concern when one is considering the theory of the firm, in which a discounted stream of profits provides economic incentive, for this is indeed an integral. But as soon as one looks at the other side of the market, one is aware that consumers are not in general likely to be maximising discounted streams of utility. Rather, it is usual to suppose that enjoyment today and enjoyment tomorrow are strongly complementary for most people. Theories of job choice, attempts to explain age-earnings profiles and the like must take full account of this fact. No doubt the mathematics of maximising general functionals will be shortly worked out well enough for economists to use it, but in the meantime the best way out seems to be to give up a description of events in continuous time and concentrate instead on objective functions that depend on a large, but finite, number of discrete variables. An inspiring example of this is to be found in a paper by Iwai (1972) in which a programme of optimal capital accumulation is worked out for a quite general benefit function in a model with a discrete time variable.

In this thesis, continuous mathematics has been used throughout, not only for time, but even for households and buildings, which must in the nature of things come in integral amounts. Approximations are involved in such a procedure as this, and it can be difficult to know when the graininess or lumpiness of people and things can make a crucial difference in an economic story. Non-convexities, like uncertainty, provoke a good deal of new economic theory these days. I am inclined to think that the models treated in the present work are not too much impaired by continuous approximations, but it is certainly reasonable to hope that intertemporal

models will be developed with discrete households, firms, and so forth, even if continuous time is (very properly!) retained.

There is no reason to imagine that aggregation over firms and households, in the sense of bringing them together in a general-equilibrium economy, is any more difficult a task with discrete variables than with continuous ones. In fact, it may well be that the *individuality* of people and managements can be captured better in a discrete model. Assumptions of perfect competition among identical agents are probably grating on the consciences of most economists nowadays, and it is certain that, if entry and exit of firms are to be objects of inquiry, one must have in mind a hierarchy of them in order of efficiency or some such attribute. An idea like this one is behind the modelling of suburban construction activity in Chapter IIIA.

It is interesting to wonder if one might construct a general-equilibrium model, with many goods and services, in which all the economic actors had intertemporally defined objectives. The correspondence discovered in Chapter I between lines of constant state or co-state variables and static supply or demand curves may well be subject to considerable generalisation. It is tantalising to imagine that the notions already in use for demand analysis in a static framework - Hicks aggregation, gross substitutability, and so forth - might have dynamic counterparts that could help to cut a way through the difficulties of many-stock optimal control problems.

I should like to end this thesis by lamenting the fact that our understanding of intertemporal economic processes has not at all

contributed to a credible theory of speculation. Well-behaved, responsible firms perhaps, but crass speculators no. They might just as well be necromancers or augurers for all we understand them - and some are indeed very successful. The glamorous mathematical theory of catastrophes (see, for a serious account of the theory, Bröcker (1975)) may provide what is needed here - it does discuss motions described by systems of differential equations and the quirks or singularities associated with such motions. I have the uneasy feeling, though, that there is still some distance to go before the economics of speculation is understood.

BIBLIOGRAPHY.

1. K.J. ARROW (1968) "Applications of Control Theory to Economic Growth" in *Lectures in Applied Mathematics*, American Mathematical Society, Providence, R.I.

2. K.J. ARROW (1971) *Essays on the Theory of Risk Bearing*, Markham.

3. K.J. ARROW & F.H. HAHN (1971) *General Competitive Analysis*, Holden-Day. (chapter 2)

4. R. BELLMAN (1961) *Adaptive Control Processes: a Guided Tour*, Princeton. (p. 56)

5. R. BELLMAN (1967) *Introduction to the Mathematical Theory of Control Processes*, Academic Press.

6. W. BROCK (1972) "On Models of Expectations that Arise from Maximising Behaviour of Economic Agents over Time", *Journal of Economic Theory*, 5, pp 348 - 376.

7. W. BROCK (1976) "The Global Asymptotic Stability of Optimal Control: A Survey of Recent Results", Report 7604, Center for Mathematical Studies in Business and Economics, University of Chicago, Jan. 1976.

8. W. BROCK & E. BURMEISTER (1976) "Regular Economies and Conditions for Uniqueness of Steady States in Optimal Multisector Economic Models", *International Economic Review*, 17, pp 105 - 120.

9. W. BROCK & J.A. SCHEINKMAN (1976) "The Global Asymptotic Theory of Optimal Control with Applications to Dynamic Economic Theory", to appear in *Applications of Control Theory to Economic Analysis*, eds. J. Pitchford and S. Turnovsky.

10. T. BRÖCKER (1975) *Differentiable Germs and Catastrophes*, (tr. L. Lander) Cambridge.

11. A.E. BRYSON Jr. & Y.-C. HO (1969) *Applied Optimal Control*, Blaisdell.

12. A.E. BRYSON Jr. & Y.-C. HO (1969a) *ibid.*, Appendix 4.

13. A.E. BRYSON Jr. & Y.-C. HO (1969b) *ibid.*, p. 110.

14. C.W. CLARK (1974) "Optimal Control Theory and Renewable Resource Management" in *Optimal Control Theory and its Applications*, ed. B.J. Kirby. Springer-Verlag.

15. W.E. DIEWERT (1974) "Intertemporal Consumer Theory and the Demand for Durables", *Econometrica* 42, pp. 497 - 516.

16. E. DOMAR (1946) "Capital Expansion, Rate of Growth and Employment", *Econometrica*, 14, pp. 137 - 147.

17. R. DORFMAN, P.A. SAMUELSON & R.M. SOLOW (1958) *Linear Programming and Economic Analysis*, McGraw-Hill (chapter 12)

18. L. EPSTEIN (1977) "The Le Chatelier Principle in Optimal Control Problems", manuscript, University of British Columbia.

19. S. GOLDMAN (1968) "Optimal Growth and Continual Planning Revision", *Review of Economic Studies*, XXXV(2), pp. 145 - 154.

20. W.M. GORMAN (1968) "Measuring the Quantities of Fixed Factors", in *Value, Capital and Growth*, ed. J.N. Wolfe, Aldine, Edinburgh.

21. J.P. GOULD (1968) "Adjustment Costs in the Theory of Investment of the Firm", *Review of Economic Studies*, XXXV(1), pp. 47 - 56.

22. G. HADLEY & M.C. KEMP (1971) *Variational Methods in Economics*, North Holland.

23. F.H. HAHN (1968) "On Warranted Growth Paths", *Review of Economic Studies*, XXXV(2), pp. 175 - 184.

24. G.H. HARDY (1952) *A Course of Pure Mathematics*, 10th edition, Cambridge.

25. R.F. HARROD (1939) "An Essay in Dynamic Theory", *Economic Journal*, 49, pp. 14 - 33.

26. T. HATTA (1977) "A Theory of Piecemeal Policy Recommendations", *Review of Economic Studies*, XLIV(1), pp. 1 - 22.

27. K. IWAI (1972) "Optimal Economic Growth and Stationary Ordinal Utility: a Fisherian Approach", *Journal of Economic Theory*, 5, pp. 121 - 151.

28. D.L. KLEIDER, R.G. KULLER & D.R. OSTBERG (1968) *Elementary Differential Equations*, Addison-Wesley. (chapter 9)

29. M.A. KRASNOSEL'SKII (1964) *Topological Methods in the Theory of Nonlinear Integral Equations*, Oxford. (p. 141 *et seq.*)

30. M. KURZ (1968) "The General Instability of a Class of Competitive Growth Processes", *Review of Economic Studies*, XXXV(2) pp. 155 - 174.

31. E.B. LEE & L. MARKUS (1967) *Foundations of Optimal Control Theory*, Wiley.

32. R.E. LUCAS (1967) "Adjustment Costs and the Theory of Supply", *Journal of Political Economy* 75, pp. 321 - 334.

33. D.T. MORTENSEN (1970) "A Theory of Wage and Employment Dynamics", in *Microeconomic Foundations of Employment and Inflation Theory*, ed. E.S. Phelps, Norton.

34. J.F. MUTH (1961) "Rational Expectations and the Theory of Price Movements", *Econometrica* 29, pp. 315 - 335.

35. K. NAGATANI (1976) "On the Comparative Statics of Temporary Equilibrium", University of British Columbia Department of Economics Discussion Paper 76-30.

36. J. von NEUMANN (1945) "A Model of General Economic Equilibrium", *Review of Economic Studies*, XIII(1), pp. 1 - 9.

37. J. von NEUMANN & O. MORGENSTERN (1944) *Theory of Games and Economic Behaviour*, Princeton.

38. H. ONIKI (1973) "Comparative Dynamics (Sensitivity Analysis) in Optimal Control Theory", *Journal of Economic Theory*, 6, pp. 265 - 283.

39. E.S. PHELPS (1961) "The Golden Rule of Accumulation: a Fable for Growthmen", *American Economic Review*, 51, pp. 638 - 643.

40. L.S. PONTRYAGIN, V.G. BOLTYANSKII, R.V. GRAMKRELIDZE & E.F. MISCHENKO (1962) *The Mathematical Theory of Optimal Processes*, Interscience.

41. R. RADNER (1961) "Paths of Economic Growth that are Optimal with Regard only to Final States: a Turnpike Theorem", *Review of Economic Studies*, XXVIII, pp. 98 - 104.

42. F.P. RAMSEY (1928) "A Mathematical Theory of Savings", *Economic Journal*, 38, pp. 543 -559. Reprinted in *Growth Economics*, ed. A.K.Sen. Penguin.

43. P.A. SAMUELSON (1947) *Foundations of Economic Analysis*, Harvard.

44. P.A. SAMUELSON (1947a) *ibid.*, pp 30 -33.

45. R.M. SOLOW (1956) "A Contribution to the Theory of Economic Growth", *Quarterly Journal of Economics*, 70, pp. 65 - 94.