

THE UNIVERSITY *of York*

CENTRE FOR HEALTH ECONOMICS

**MAKING STATIC MODELS DYNAMIC:  
THE CASE OF THE NATIONAL HEALTH  
SERVICE**

Ann van Ackere  
Peter Smith

*CHE Technical Paper Series 9*



# **MAKING STATIC MODELS DYNAMIC: THE CASE OF THE NATIONAL HEALTH SERVICE**

**Ann van Ackere (London Business School) and  
Peter C. Smith (University of York)**

**October 1997**

## **CENTRE FOR HEALTH ECONOMICS TECHNICAL PAPER SERIES**

The Centre for Health Economics has a well established Discussion Paper series which was originally conceived as a means of circulating ideas for discussion and debate amongst a wide readership that included health economists as well as those working within the NHS and pharmaceutical industry.

The introduction of a Technical Paper Series offers a further means by which the Centre's research can be disseminated. The Technical Paper Series publishes papers that are likely to be of specific interest to a relatively specialist audience, for example papers dealing with complex issues that assume high levels of prior knowledge, or those that make extensive use of sophisticated mathematical or statistical techniques.

The series provides a maximum of 20 pages for authors. The content and its format are entirely the responsibility of the author, and papers published in the Technical Paper series are not subject to peer-review or editorial control, unlike those produced in the Discussion Paper series. Offers of further papers, and requests for information should be directed to Frances Sharp in the Publications Office, Centre for Health Economics, University of York.

### **ACKNOWLEDGEMENTS**

The authors have benefited from the helpful comments of Hugh Gravelle, Stephen Martin and other colleagues at the University of York, of Erik Larsen at the University of Bologna and from students on the course "Systems Dynamics Applications and Advanced Topics" in the Department of Information Science, University of Bergen.

Address for correspondence:

Peter C. Smith  
Department of Economics & Related Studies  
University of York  
York YO1 5DD  
England

Phone: + 44 1904 433779  
Fax: + 44 1904 433759  
E-mail: [pcsl@york.ac.uk](mailto:pcsl@york.ac.uk)

Price £4.00

©Ann van Ackere and Peter C. Smith

## ABSTRACT

Equilibrium is probably the principal focus of most areas of economic analysis. However, the thesis of this paper is that the policy maker is often interested not only in the equilibrium predictions arising from an economic model, but also in the path taken by policy variables as they move towards that equilibrium. It is argued that integration into a dynamic framework is likely to enhance the usefulness of an economic model whenever this is the case. The principles are illustrated using an example from the British National Health Service, in which a traditional economic model of supply and demand is deployed within a dynamic systems model.

**Keywords:** methodology; dynamics; policy; health care.

## INTRODUCTION

In two important respects almost all social systems are dynamic in nature. First, the components of the system are constantly changing, as new technologies are developed, new information is revealed, and individual preferences and wealth change. And second, the system can move only gradually towards any implied equilibrium - adjustment is rarely instantaneous. As a result, at any moment in time, rather than being at rest, a social system is likely to be some distance away from an equilibrium. Indeed, it is questionable whether equilibrium is in any sense a "natural" state (Milgate, 1987).

Yet economists have traditionally been interested more in the equilibrium than in the path whereby it is reached. Marshall (1890) set the tone for the development of the discipline of economics by conceiving equilibrium as the state of affairs that arises under "long-run normal conditions". A large part of the research effort of economists has since been devoted to establishing the existence and exploring the nature of such equilibria. Interest in situations of disequilibrium have tended to be restricted to the macro-economics field, starting with the work of Keynes, whose views on "the long run" are well known.<sup>1</sup> In this tradition, disequilibrium is assumed to arise from the existence of various frictions or rigidities in the economy, such as "sticky" prices, institutional barriers, and the failure of agents to respond to economic signals.

Recognition of the existence of such rigidities has given rise to the concept of the temporary equilibrium (Grandmont, 1987). Recognising the central role of expectations in the analysis of equilibrium, this line of economic enquiry focuses on how agents' expectations are formed. For example, expectations may be adaptive, in the sense that the expectations can be thought of as a stock which is only gradually altered by the flow of current information. Models such as this imply that agents are in some sense "slow" to learn. This implies that there is a delay between when information becomes available, and when it is fully incorporated into the decision making process. The economy may then evolve via a series of temporary equilibria, which may or may not converge to a long run equilibrium. In the spirit of the temporary equilibrium, Hicks (1985) conceived the notion of the "traverse", which sought to indicate the optimal path from one equilibrium to another.

Nevertheless, most of these approaches still assume the existence of a desirable equilibrium, which will somehow be reached. Arthur (1994) looks at the issue of *path dependence* - that is, the extent to which the final outcome depends not only on actions taken now, but also on what has happened in the past. He demonstrates how small, seemingly unimportant events may have a considerable impact on the eventual outcome. That is, the "obvious" or "most desirable" equilibrium may not be reached due to some events early on the path of development. Competition between two alternative technologies can be used as an illustration. There are several examples where an inferior technology has succeeded in driving a superior technology out of the market as a result of certain events taking place early in the life of these products (such as, for example, a better marketing campaign leading to higher initial sales, thereby generating increased funds to create market dominance for the inferior product). Frequently cited examples include video tapes (VHS against Betamax) and computers (IBM compatibles against Apple): in both cases an apparently better technology failed to establish itself as the market leader. Another interesting example is that of the QWERTY keyboard layout: it is

generally accepted that this layout is inefficient (it was initially designed to minimise jamming on mechanical typewriters), but there is great reluctance to switch to a more effective layout, despite the general agreement that it would take an experienced typist only about 10 days to retrain, and that the benefits would by far outweigh the costs.

Policy makers are clearly interested in equilibria in the sense that they illustrate the expected long run consequences of policy decisions, and economists have often been influential in informing policy on the basis of equilibrium analysis. Recent examples in the United Kingdom have been the 1990 reform of local government finance (Butler, Adonis and Travers, 1994) and the 1991 reform of the National Health Service (Robinson and Le Grand, 1994). However, to the policy maker, the short run is often of central importance. Therefore, not only is a long run equilibrium of interest - so is the *path* which is followed to reach equilibrium. Indeed, given the short time horizon of policy makers, the early stages of the path may be much more important than the equilibrium, which may never be reached due to frequent policy changes and environmental uncertainty. In this respect - other than the specialist field of macro-economic forecasting - economics has had relatively little to say to policy makers.

The purpose of this paper is to indicate how conventional micro-economic models can be adapted to offer guidance to policy makers on the dynamic implications of policy initiatives. The key insight is that social systems are characterised by rigidities, in the sense that some elements cannot be changed instantly. These rigidities can be conceived as stocks - something that accumulates or depletes only gradually. In any time period these stocks are augmented or diminished by their respective flows. The rate of accumulation and depletion depends on the flow rates during that period. The flows in turn are triggered by information and the levels of the stocks. Therefore, the level of a stock responds only gradually to new flows, so that in general the ramifications of a major shock to the system unfold only over a number of time periods. Mathematically, of course, such relationships can be modelled using systems of ordinary differential or difference equations. However, in this paper we propose to use methods that are less opaque to the general reader.

The stocks to which we refer can take a number of forms. They may be physical, for example in the form of physical capital. They may be socially constructed, for example in the form of long term contracts or other commitments. Or they may be perceptual, taking the form of preferences or beliefs. The feature common to all these phenomena is that they cannot change instantaneously in response to shocks. Central to the paper is therefore the concept of partial adjustment, as found in numerous situations where instantaneous full adjustment to new circumstances is not possible.

The second central foundation of the paper is the requirement that the policy variables of interest are embedded within a system of feedback. That is, they are determined in part by endogenous variables, which have themselves in turn been influenced, either directly or indirectly, by the values assumed by the policy variable at some time in the past. Because of data limitations, such endogeneity is often modelled in economics using systems of simultaneous equations, which seek to capture the "simultaneous" determination of endogenous variables. This approach is useful if the interest is in equilibrium. It is however less relevant if dynamic issues are of interest. Relationships are only truly "simultaneous" if equilibrium has

been attained. Otherwise, complex cause and effect relationships between variables are likely to be observed. Given our interest in the paths taken by policy variables and the resulting system behaviour, it becomes necessary to model such relationships explicitly.

The methods we propose entail the integration of conventional micro-economic models with the methods of system dynamics. The system dynamics approach encompasses the two key premises discussed above: first, that the structure of a system drives its behaviour, which in turn drives the events; and second, that structure can be represented by a number of interacting feedback loops. The focus is on understanding how the underlying structure drives behaviour so as to understand how events unfold over time. We shall illustrate these principles using an example from the UK National Health Service.

## SYSTEMS DYNAMICS

System dynamics dates back to the late fifties, and interest in the methodology grew rapidly during the sixties and early seventies. The initial focus was on the application of system dynamics management issues, but soon included the analysis of social and macro-economic problems. See Forrester (1961, 1968, 1971) for details of this early work. Roberts (1984) contains a collection of early papers. Since the mid-eighties, there has been renewed interest in applying system dynamics to business policy and strategy problems. This interest has been facilitated by the availability of new, user friendly, high level graphical simulation programs (such as *ithink*, *Powersim* and *Vensim*<sup>2</sup>). Easily accessible books describing the system dynamics approach (for example, Senge (1991); Morecroft and Sterman (1994)) have also played a key role. Previous applications in health care include Ittig (1976), Wolstenholme (1993), and Vennix and Gubbels (1994). The model presented in this paper was build using *ithink*.

The structure of a system can be sketched using causal loops. Causal loops can be either balancing (capturing negative feedback) or reinforcing (capturing positive feedback). A balancing loop in isolation exhibits a goal seeking behaviour: that is, after a disturbance, the system seeks to return to an equilibrium situation (conforming to the notion of a stable equilibrium). A reinforcing loop in isolation exhibits exponential growth or decay results: that is, an initial disturbance leads to further change, suggesting the presence of an unstable equilibrium. Understanding the behaviour of one loop in isolation is straightforward. Unfortunately, the human brain is poor when it comes to predicting the behaviour resulting from interacting feedback loops. This creates the need to move from causal loop diagrams to more formal models which can be simulated.

A system dynamics simulation model consists of two components: the stock and flow network, and the information network. Stocks represent elements of the model which cannot be changed instantly, they accumulate or deplete gradually, regulated by their in- and out-flows. Stocks can be “hard” (tangible, easily measurable) concepts, such as the number of beds for elective surgery, or “soft” concepts (such as the average waiting time as perceived by patients). The rates of the flows are determined by the information network, and depend on the level of the various stocks in the system. These rates can be interpreted as the output of policies, or decision making processes. For instance, one stock could represent the number of beds allocated to

elective surgery. The in- and out-flows represent increases and decreases in this allocation. These rates of change result from a decision process, based on various pieces of information, such as the present number of beds, the waiting time, the targeted waiting time, needs for other types of care, and so on.

## **ELECTIVE SURGERY IN THE NHS**

The UK National Health Service (NHS) is a massive public sector organisation funded out of general taxation. It seeks to deliver all aspects of health care to UK citizens free of charge at the point of delivery. It represents about 5.4% of gross domestic product, and employs almost a million employees. Although generally considered a highly efficient organisation in terms of health gain in relation to expenditure, the NHS has for many decades suffered from massive waiting lists for elective surgery, and very long waiting times for many patients (Yates, 1987). Historically, policy emphasis had been on the size of the waiting *lists*. However, as part of its Citizen's Charter initiative, the UK Government (1992) introduced a requirement that no patient should have to wait longer than 24 months for elective surgery, signalling a change of focus to waiting *times*. The limit has subsequently been reduced to 18 months for all patients, and less for certain procedures (UK Government, 1995). Various aspects of waiting times at individual NHS hospitals are published in annual performance guides which receive widespread political and media attention.

Since 1991 the NHS has been organised as an internal market, in which purchasers (health authorities or general practitioner fundholders) negotiate contracts for health care with providers (such as NHS hospitals). The providers therefore compete for business from NHS purchasers, and usually depend for their continued existence on the NHS contracts they succeed in negotiating. In the light of the Patient's Charter, purchasers and general practitioners are likely to be influenced by waiting times in their choice of providers, and indeed waiting times have become a central issue in many contracts. In particular, many general practitioner fundholders insist on certain maximum waiting times for certain elective procedures. NHS patients can only receive elective surgery at an NHS provider if they are referred to that provider by a general practitioner. Therefore managers and physicians within providing organisations are in turn likely to pay a great deal of attention to waiting times.

## **A SYSTEMS DYNAMICS MODEL OF WAITING TIME IN THE NHS**

We model a highly simplified version of the real system outlined above, our aim being to capture the key dynamics without any unnecessary detail. In other words, we look for the simplest possible model which incorporates the basic feedback structure described in Figure 1. Our approach can be thought of as a 'model of models'. We therefore use a highly aggregate system dynamics model to represent the key feedback characteristics, in the belief that further work can elaborate the model if it is found to be useful in this initial form.

Figure 1 shows a simple causal loop diagram illustrating the key feedback structure of average waiting time for National Health Service (NHS) elective surgery, which consists of two balancing loops, as indicated by the B in the centre of the loops. The left loop represents the demand side: other things being equal, higher demand leads to longer average waiting times

(demand and waiting time move in the same direction, as indicated by the S on the arrow), and longer average waiting times lead to lower demand (waiting time and demand move in opposite direction, as indicated by the O on the arrow). The double bar // indicates a delay: it takes time for patients (and/or their general practitioner) to perceive a change in average waiting time and adapt their behaviour by looking for alternative forms of treatment. The right loop represents the supply side: increased waiting times result, after some delay, in more resources being allocated to elective surgery (as indicated by the S on the arrow), which leads to shorter average waiting times (indicated by the O on the arrow). This is a highly simplified representation, and the more detailed model will be presented later. Still, this simple model will enable us to illustrate the insights which can be gained from focusing on the transition path from one equilibrium to another.

We base our dynamic model on an econometric study undertaken by Martin and Smith (1995) for the UK Department of Health, which was itself based on the simple static equilibrium models developed by Lindsay and Feigenbaum (1984), Cullis and Jones (1986) and Goddard, Malek and Tavakoli (1995). This literature recognises that waiting time acts as a "price" for patients seeking health care free at the point of consumption. The Martin and Smith study examined the utilisation of facilities for routine non-emergency surgery in the UK National Health Service (NHS). Demand for surgery was assumed to depend on local waiting times, clinical need, the provision of family practitioner services and the availability of private health care, as follows (with predicted influences in parentheses):

$$\begin{aligned} \text{Utilization}_{\text{demand}} = f &(\text{waiting time } (-), \\ &\text{need } (+), \\ &\text{family practitioner supply } (?), \\ &\text{provision of private inpatient beds } (-)) \end{aligned}$$

The model of the supply of surgical resources within the NHS incorporated the total NHS budget available to local managers as well as local waiting time. The empirical supply equation is:

$$\begin{aligned} \text{Utilization}_{\text{supply}} = g &(\text{waiting time } (+), \\ &\text{provision of NHS beds } (+)). \end{aligned}$$

Note that in this model, the provision of NHS beds is used as a measure of the total acute sector resources available to local NHS managers, comprising capital, labour and support services, as well as the beds themselves. Clearly the acute sector budget would be a better measure in this respect, but accurate budget data are not systematically available within the NHS, and we believe that bed availability is likely to be a satisfactory proxy.

Using conventional econometric methodology, Smith and Martin assumed the above system was in equilibrium, and estimated the model empirically using cross-sectional data from over 4,000 small areas in England. The result was the estimation of a system of equations, of which the most important were the following:

WAIT_TIME =	$-0.071 + 0.203 \text{ NEED} - 0.024 \text{ FP_SUPPLY} - 0.241 \text{ NHS_BEDS}$
ELECTIVES =	$-.860 + 0.291 \text{ WAIT_TIME} + 0.211 \text{ NHS_BEDS} + 0.148 \text{ DAY_CASES} - 0.240 \text{ LENGTH_STAY} - 0.056 \text{ FP_SUPPLY} - 0.232 \text{ PRIV_BEDS} - 0.196 \text{ NEED} + 0.120 \text{ HOMES}^*$
UTILIZATION =	$-1.539 - 0.089 \text{ WAIT_TIME} - 0.107 \text{ FP_SUPPLY} - 0.091 \text{ PRIV_BEDS} + 0.800 \text{ NEED}$

where

WAIT_TIME	Standardized waiting times for elective surgery, 1991-92
UTILIZATION	Standardized utilisation rate (episodes), 1991-92
NHS_BEDS	Provision of NHS beds
FP_SUPPLY	Family practitioner supply
NEED	Weighted index of health needs
ELECTIVES	Proportion of all episodes that are elective
DAY_CASES	Proportion of all elective episodes that are day cases
LENGTH_STAY	Standardized (for age and sex) length of stay
PRIV_BEDS	Accessibility of private hospital beds
HOMES*	Proportion of residents aged 75+ NOT in residential/nursing homes

Logarithms were taken of all variables, so the coefficients can be interpreted as elasticities.

Solving these equations suggested that a (permanent) increase in NHS resources (NHS\_BEDS) would eventually result in substantial reductions in NHS waiting times without stimulating a large concomitant increase in demand. The study therefore proved useful in the budget negotiations between the UK Treasury and the Department of Health, as it gave quantitative guidance on the long run equilibrium implications for waiting times and resource use of an increase in the NHS budget.

However, notwithstanding the policy relevance of these results, the study begged a number of questions which the econometric analysis was ill-equipped to address. The most pressing of these was: bearing in mind the manifest rigidities in the NHS system, how long would it take for the impact of increased resources to be fully reflected in reduced waiting time? Moreover, numerous subsidiary questions arose. For example, what would be the long term impact of increased NHS resources on the extent of private health insurance coverage? how would the increase affect the non-surgical part of the NHS? to what extent would the increase affect NHS efficiency levels? and what would be the long run impact on the market for private health care?

Clearly these concerns (and many other associated issues) are linked within a complex dynamic system of cause and effect. The conventional economic approach is to model such links using a system of simultaneous equations. In contrast, the system dynamics methodology seeks to model each of these links explicitly, and follow the resulting behaviour over time. Thus to parametrize our dynamic model, we use the outputs of more traditional static econometric methods. This raises the issue of whether these estimates (obtained under the assumption of equilibrium conditions) are valid inputs for a dynamic model. Our claim is, that while the use

of such data may not be a perfect solution, it is a significant improvement over the assumption of (say) a constant elasticity. We make the point that, as the system moves from one 'state' to another (i.e. as the length of the waiting list and the level of resources change), the strength of the various feedbacks changes. We use estimates obtained for each of these various states as proxies for the true elasticities. It is also worth noting that our aim is not to build a forecasting model. Rather, the purpose of this exercise is to gain insight into these feedback mechanisms. Therefore, it is the relative magnitude of the elasticities in the different states which is relevant, rather than the precise numerical values used.

Referring back to Figure 1, we used NHS beds allocated to elective surgery as a proxy for resources. The stock, flow and information network is shown in Figure 2. The equations (of which the most important are discussed below) are listed in the appendix. The model contains five stocks, indicated by rectangles. Three of these (number of people on the waiting list, number of beds and expressed demand) are hard, measurable quantities. In the remainder of this paper, "demand" refers to expressed demand, unless mentioned otherwise. The remaining two (waiting time as perceived by the patient and general practitioner on the demand side, and waiting time as perceived by hospital management on the supply side) represent perceptions, i.e. they attempt to capture how the two main actors adjust their perception of average waiting time over time as new information becomes available. These are modelled separately, as hospital management has better access to relevant, accurate, up to date information, and is thus able to form a more accurate judgement over a shorter period of time. As discussed below, this difference is captured by the variables called 'time to perceive waiting time'.

The stocks act as rigidities, in the sense that they do not change instantly, but only as the result of in- and out-flows. These are represented by double arrows. The white head represents the direction of flow. For instance, when 'change in demand' is positive, demand increases, while if 'change in demand' is negative, demand decreases. Note that some flows are uniflows, e.g. 'patients treated' is always non-negative.

To estimate stock levels, one must therefore explicitly model the rates of their in- and out-flows. In this respect it is worth emphasising the difference between actual average waiting time, and patients' perception of waiting time. We assume that decisions are based on what people perceive waiting times to be, rather than on actual values, unknown to them. A nice illustration of the difference between perception and reality is provided in Barnett and Saponaro (1985) in the context of users of the Boston underground. They show empirically that (i) users of the Red line significantly over-estimate the frequency of lengthy waiting times (ii) they do so to a much larger extent than users of the Blue line over the same period. Barnett and Saponaro argue that this is a consequence of low service standards in a previous period, where lengthy waiting times were the norm. That is, passengers' present perception of waiting times are heavily influenced by past experience. Information links are represented by arrows in the diagram. For instance, 'waiting time' is determined by the values of the stock 'waiting list' and the flow 'patients treated'.

Next we describe the main relationships in the model. The stock 'Waiting list' is replenished by the flow 'Referrals' and depleted by the flow 'Patients treated'. The rate of referrals is determined by the level of expressed demand (calculated as patients per month). The number of

patients treated per month depends on the number of beds, and the average length of stay. The stocks 'Demand' and 'Beds' are affected by the flows 'Change in Demand' and 'Change in Beds' respectively. The change in beds is driven by the elasticity of beds with respect to average waiting time as perceived by the supply side. The Martin and Smith (1995) study yields an elasticity estimate of 0.29, implying a 0.29% increase in beds for a 1% increase in waiting time. This reflects the internal pressures on allocation of resources between elective surgery and other forms of care. Appendix 1 lists the illustrative equations we have chosen for this demonstration of the full model, using a mixture of evidence and judgement.

Change in demand is driven by the elasticity of demand with respect to average waiting time as perceived by the demand side. The original Martin and Smith (1995) study yielded an average estimate of -0.09. However, further analysis of the data indicated that this value varies significantly as a function of average waiting time. We therefore represented elasticity of demand by a graphical function as shown in Figure 3 (a). The estimates indicate that the elasticity is not significantly different from zero for average waiting times up to about 3 months. It then decreases quite sharply, to reach a value of about -4.0 for waiting times of 4 to 5 months. Estimates for longer waiting times are less reliable due to the scarcity of data, but indications are that the value returns to zero. This implies a demand function as sketched in Figure 3 (b).

This demand function seems to indicate the existence of some "acceptable" level of waiting time (about 3 to 4 months according to these estimates) which has little influence on patient demand. Once this threshold is exceeded, however, demand drops sharply. This can be interpreted as patients and their general practitioners looking for alternative forms of treatment (including private care), or patients electing not to receive treatment (see Goddard, Malek and Tavakoli, 1995). Further work might seek to model this process explicitly.

The perceived waiting time is modelled as a process of adjustment, where the perceived value is gradually brought in line with the actual value. We model the 'Change in perceived waiting time' as being equal to a fraction of the difference between the 'Perceived waiting time' and the 'Waiting time', the inverse of the fraction being labelled 'Time to perceive waiting time'. Mathematically, the perceived waiting time is a simple smoothed average of the waiting time with smoothing constant  $1/\text{Time to perceive waiting time}$ . The model assumes that hospital management makes decisions based upon a much more accurate and up to date perception of average waiting time than the demand side (patients and general practitioners). We therefore set the 'Time to average waiting time' equal to one month for the supply side, and 12 months for the demand side.

Note that we assume that both decision makers react immediately to changes in the perceived average waiting time. While this may be accurate for the supply side, it probably underestimates the delays on the demand side, where decisions such as acquiring or cancelling private insurance take time to implement.

## SOME ILLUSTRATIVE RESULTS

Figure 4 illustrates what insight can be gained from simulating this type of model. We consider four scenarios, as described below and summarized in Table 1.

	Initial waiting time	Resource change in month 10
Scenario 1	3 months	10% increase
Scenario 2	4.5 months	10% increase
Scenario 3	3 months	10% decrease
Scenario 4	4.5 months	10% decrease

**Table 1. Overview of scenarios**

We consider two initial equilibria, with respectively a 3 month and 4.5 month average waiting time. For each equilibrium we consider the impact of two alternative shocks taking place in month 10 of the simulation: a 10% increase in NHS resources and a 10% decrease. These are modelled using the ‘External change in beds’ flow. All simulations are run for 60 months. Figure 4 illustrates the dynamic impact of these shocks on four variables: average waiting time, waiting list, NHS beds devoted to elective surgery and demand. We assume unchanging efficiency, as reflected in the ‘Average length of stay’, so the number of patients treated is a constant multiple of the number of beds.

Scenarios 1 and 3 have an initial average waiting time of 3 months, implying a low elasticity of demand. For these scenarios, the results are very much as one would expect. An increase in resources (scenario 1) leads to a gradual decrease in both average waiting time and waiting list (Figures 4 (a) and (b)). There is no impact on demand, as demand elasticity is zero in this region (Figure 4(d)). Note (Figure 4 (c)) that the additional resources are gradually diverted from elective surgery to other purposes.

A decrease in resources (Scenario 3) results initially in longer waiting times and lists (Figures 4(a) and (b)). This creates pressure to re-allocate beds to elective surgery (figure 4(c)) and, with some delay, a decrease in demand is observed (Figure 4(d)). This in turn leads to a reduction of waiting time and waiting lists starting around month 22. After some further fluctuations, the system stabilises at a new equilibrium with somewhat longer waiting time and waiting list, and slightly lower demand (implying that some of the demand has been suppressed).

Scenarios 2 and 4 have an initial waiting time of 4.5 months, which situates them initially in the area of high demand elasticity. First consider scenario 2, an increase in resources. The sudden increase in resources results in an immediate shortening of the waiting time and list (Figures 4(a) and (b)). The decrease in waiting time leads to some of the additional beds being reallocated to other areas (Figure 4(c)), but demand increases sharply (Figure 4(d)). This causes waiting times and waiting lists to increase, and the lost beds reverting back to elective surgery.

About a year after the increase in resources (around month 24) the picture looks bleak: waiting lists reach a peak of 500 (i.e. a more than 10% increase, Figure 4(a)) and waiting time is slightly higher than at the start of the simulation (Figure 4(b)). Demand, having reached a high of about 115 (a 15% increase, Figure 4(c)), is on a downward trend. This implies that we are treating significantly more people than we used to, but they have to wait longer. As the dust settles (around month 35, i.e. about 2 years after the change) demand is approximately 10% above the initial level (i.e. part of the suppressed demand has surfaced), elective surgery has managed to hold on to its increased resources, waiting lists are longer by about 5% (despite the increase in

resources!), while the average waiting time is marginally shorter.

It is worth emphasising that in this scenario, one year after the resource increase, the situation actually looks worse than it was initially, both in terms of waiting time and waiting list. From a political point of view, this may be a highly undesirable situation. While the resulting equilibrium is attractive (more people treated and lower average waiting time), this "worse before better" path may be problematic.

Next consider scenario 4, a decrease in resources. The sudden decrease in resources results almost in a mirror image of scenario 2: the increase in waiting time (and list) leads to a significant reduction in demand (i.e. demand is being suppressed), while elective surgery is able to recover some of its lost resources from other departments (figure 4(c), beds increase for about 3 months). But the decrease in demand pushes waiting lists and time down, and so the regained resources are lost. When the dust settles, demand is about 10% below the initial level, elective surgery has been unable to make up for the lost resources, waiting lists are about 5% shorter, while waiting time is marginally longer.

## DISCUSSION

This paper has demonstrated how it is possible to embed a simple static economic model within a dynamic framework using the systems dynamics methodology. In the absence of adequate data - particularly as regards the formation of perceptions - several assumptions had to be made.

However it is a trivial matter to test alternative specifications within the system dynamics framework. Indeed its strength is that it readily permits examination of a wide range of alternative scenarios.

Our analysis emphasises the difference between waiting lists and waiting times. The model assumes that both suppliers and patients react to changes in waiting time rather than waiting lists. Long lists are likely to be politically acceptable, if this means that more people are being treated, and they face a shorter average waiting times. In a situation where people have little or no information about priorities and 'queue-jumping', people care about how long they have to wait, not about how many people are treated before they are. Therefore levels of satisfaction and resulting behaviour will be driven by waiting times. From a public welfare point of view, people's health is affected by how long they need to wait for surgery, not by the length of the list. This indicates that the appropriate performance measure is waiting times rather than waiting lists.

Another conclusion is the observation that the better the starting position, the easier it may be to achieve further improvement. Achieving significant improvement when starting from a situation with long waiting times is more difficult, as it requires dealing with a significant amount of suppressed demand, reflected by the high demand elasticity for waiting times in the range of 4 to 5.5 months. In this case, the resource increase barely affects the average waiting time, but has a significant impact on how many people are being treated and the length of the waiting list. This raises interesting issues about how to evaluate the performance of health authorities (purchasers within the NHS).

By explicitly modelling the rigidities in the system and the links between them, this model offers some insight into the dynamic implications of the economic model of demand and supply.

However, numerous extensions could be readily incorporated into the model to increase the richness of the analysis. For example, no explicit reference is made to the impact on the private sector of the policy changes. It may be desirable to examine the capacity for NHS surgery in terms of the beds provided and the efficiency with which they are used. It may be valuable to open up the "black box" labelled demand to gain an understanding of the extent to which changes in NHS referrals are due to changes in the behaviour of patients, general practitioners or surgeons. And it may be possible to incorporate technological change into the model. Such developments suggest an ambitious research agenda.

## REFERENCES

Arthur, W.B. (1994), *Increasing Returns and Path dependence in the Economy*, Ann Arbor: University of Michigan Press.

Barnett, A. and Saponaro, A. (1985), "Misapplications Reviews: The Parable of the Red Line", *Interfaces* 15(2), 33-39.

Butler, D., Adonis, A. and Travers, T. (1994), *Failure in British government. The politics of the poll tax*, Oxford: Oxford University Press.

Cullis, J. G. and Jones, P. R. (1986), "Rationing by waiting lists: an implication", *American Economic Review*, 76(1), 250-256.

Forrester, J. W. (1961), *Industrial dynamics*, Cambridge: MIT-Press.

Forrester, J.W. (1968), *Urban Dynamics*, Cambridge: MIT-Press.

Forrester, J.W. (1971) *World Dynamics*, Cambridge: MIT-Press.

Goddard, J. A., Malek, M. and Tavakoli, M. (1995), "An economic model of the market for hospital treatment for non-urgent conditions", *Health Economics*, 4(1), 41-55.

Grandmont, J. M. (ed) (1987), *Temporary equilibrium - selected readings*, New York: Academic Press.

Hicks, J. (1985), *Methods of dynamic economics*, Oxford: Oxford University Press.

Ittig, P.T. (1976), "A Model for Planning Ambulatory Health Services", *Management Science* 24(10), 1001-1010.

Lindsay, C. M. and Feigenbaum, B. (1984), "Rationing by waiting lists", *American Economic Review*, 74(3), 404-417.

Marshall, A. (1890), *Principles of economics*, London: Macmillan.

Martin, S. and Smith, P. (1995) *Modelling waiting times for elective surgery*, Occasional Paper, Centre for Health Economics, University of York.

Milgate, M. (1987), "Equilibrium: development of the concept", in Eatwell, J., Milgate, M. and Newman, P., *The new Palgrave dictionary of economics*, London: Macmillan.

Morecroft, J. D. W. and Sterman, J.D. (1994), *Modelling for learning organisations*. Portland: Productivity Press.

Roberts E. (ed) (1984), *Managerial Applications of System Dynamics*, Cambridge:MIT Press.

Robinson, R. and Le Grand, J. (1994), *Evaluating the NHS reforms*, London: King's Fund Institute.

Senge, P. (1991), *The fifth discipline: The art and practice of the learning organization*, , New York: Double Day.

Wolstenholme, E. R. (1993), "A case study in community care using systems thinking", *Journal of the Operational Research Society*, 44, 925-934

UK Government (1992), *The Citizen=s Charter*, London: HMSO.

UK Government (1995), *The Patient=s Charter and you*, London: Department of Health.

Vennix, J. and J. Gubbels (1994), *Knowledge Elicitation in Conceptual Model Building: A Case study in Modeling a Regional Dutch Health Care System*, in *Modeling for Learning Organisations*, J. Morecroft and J. Sterman (ed), Portland: Productivity Press..

Yates, J. (1987), *Why are we waiting?*, Oxford: Oxford University Press.

## Footnotes

1 "But this long run is a misleading guide to current affairs. *In the long run* we are all dead." *Tract on Monetary Reform* (1923) chapter 3.

2. *Powersim* is a registered trademark of Modeldata AS, *ithink* is a registered trademark of High Performance Inc., *Vensim* is a registered trademark of Ventana. *ithink* is available from *Cognitus* in the UK and *High Performance Systems* in the US. No prior programming experience is required. The retail price in the UK is around £800, educational discounts are available.

## APPENDIX: SPECIFICATION OF THE DYNAMIC MODEL

### Sector 1: Waiting list

$\text{Waiting\_list}(t) = \text{Waiting\_list}(t - dt) + (\text{Referrals} - \text{Patients\_treated}) * dt$   
 INIT Waiting\_list = 450 {patients}  
 INFLOWS: Referrals = Demand {patients per month}  
 OUTFLOWS: Patients\_treated = Beds/Average\_length\_of\_stay {patients per month}

$\text{Perc\_waiting\_time\_D}(t) = \text{Perc\_waiting\_time\_D}(t - dt) + (\text{Change\_in\_perc\_waiting\_time\_D}) * dt$   
 INIT Perc\_waiting\_time\_D = Waiting\_list/Patients\_treated {months}  
 INFLOWS: Change\_in\_perc\_waiting\_time\_D = (Waiting\_time - Perc\_waiting\_time\_D)/Time\_to\_perc\_waiting\_time\_D {months per month}

$\text{Perc\_waiting\_time\_S}(t) = \text{Perc\_waiting\_time\_S}(t - dt) + (\text{Change\_in\_perc\_waiting\_time\_S}) * dt$   
 INIT Perc\_waiting\_time\_S = Waiting\_list/Patients\_treated {months}

INFLOWS: Change\_in\_perc\_waiting\_time\_S = (Waiting\_time - Perc\_waiting\_time\_S)/Time\_to\_perc\_waiting\_time\_S {months per month}

Average\_length\_of\_stay = .1 {months}  
 Time\_to\_perc\_waiting\_time\_D = 12 {months}  
 Time\_to\_perc\_waiting\_time\_S = 1 {months}  
 Waiting\_time = Waiting\_list/Patients\_treated {months}

### Sector 2: Changes in drivers

$\text{Beds}(t) = \text{Beds}(t - dt) + (\text{Change\_in\_beds} + \text{External\_change\_in\_beds}) * dt$   
 INIT Beds = 10 {beds}  
 INFLOWS: Change\_in\_beds = Elasticity\_of\_beds\*Beds\*Change\_in\_perc\_waiting\_time\_S/Perc\_waiting\_time\_S {beds per month}

$\text{External\_change\_in\_beds} = \text{GRAPH}(\text{Time})$   
 {Used to model the resource changes in the various scenarios}

$\text{Demand}(t) = \text{Demand}(t - dt) + (\text{Change\_in\_demand}) * dt$   
 INIT Demand = 100 {patients per month}  
 INFLOWS: Change\_in\_demand = Elasticity\_of\_demand\*Demand \*Change\_in\_perc\_waiting\_time\_D/Perc\_waiting\_time\_D {people per month}

Elasticity\_of\_beds = .29 {constant}  
 Elasticity\_of\_demand = GRAPH(Perc\_waiting\_time\_D)  
 (0.00, 0.00), (0.5, 0.00), (1.00, 0.00), (1.50, 0.00), (2.00, 0.00), (2.50, 0.00), (3.00, 0.00), (3.50, -0.975),  
 (4.00, -4.00), (4.50, -4.00), (5.00, -4.00), (5.50, -0.975), (6.00, 0.00) {See figure 3 for a graphical representation}

Figure 1. Feed back structure of NHS elective surgery waiting times

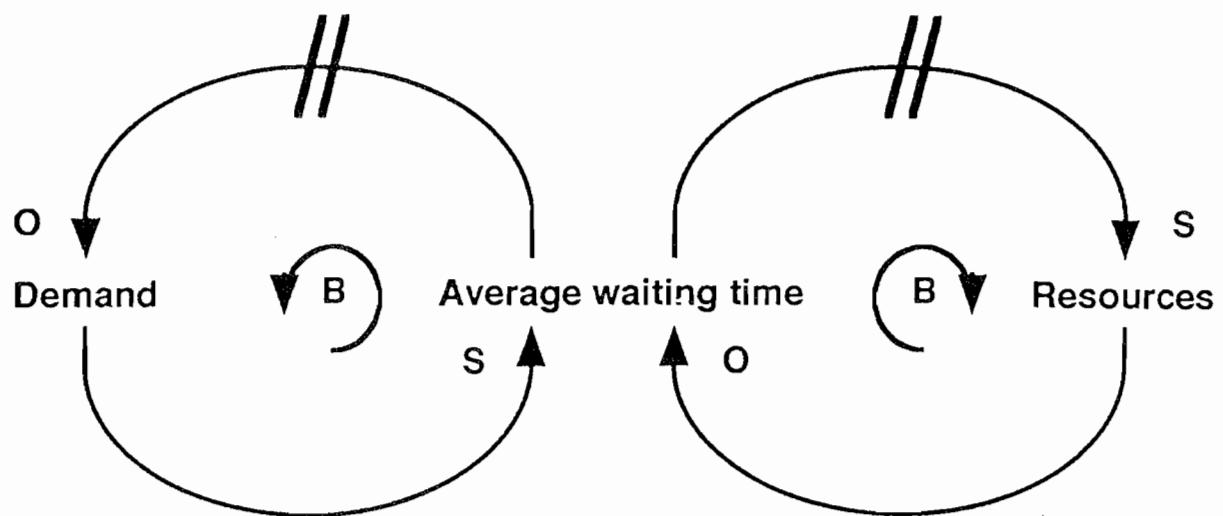


Figure 2. Stock, Flow and Information Network

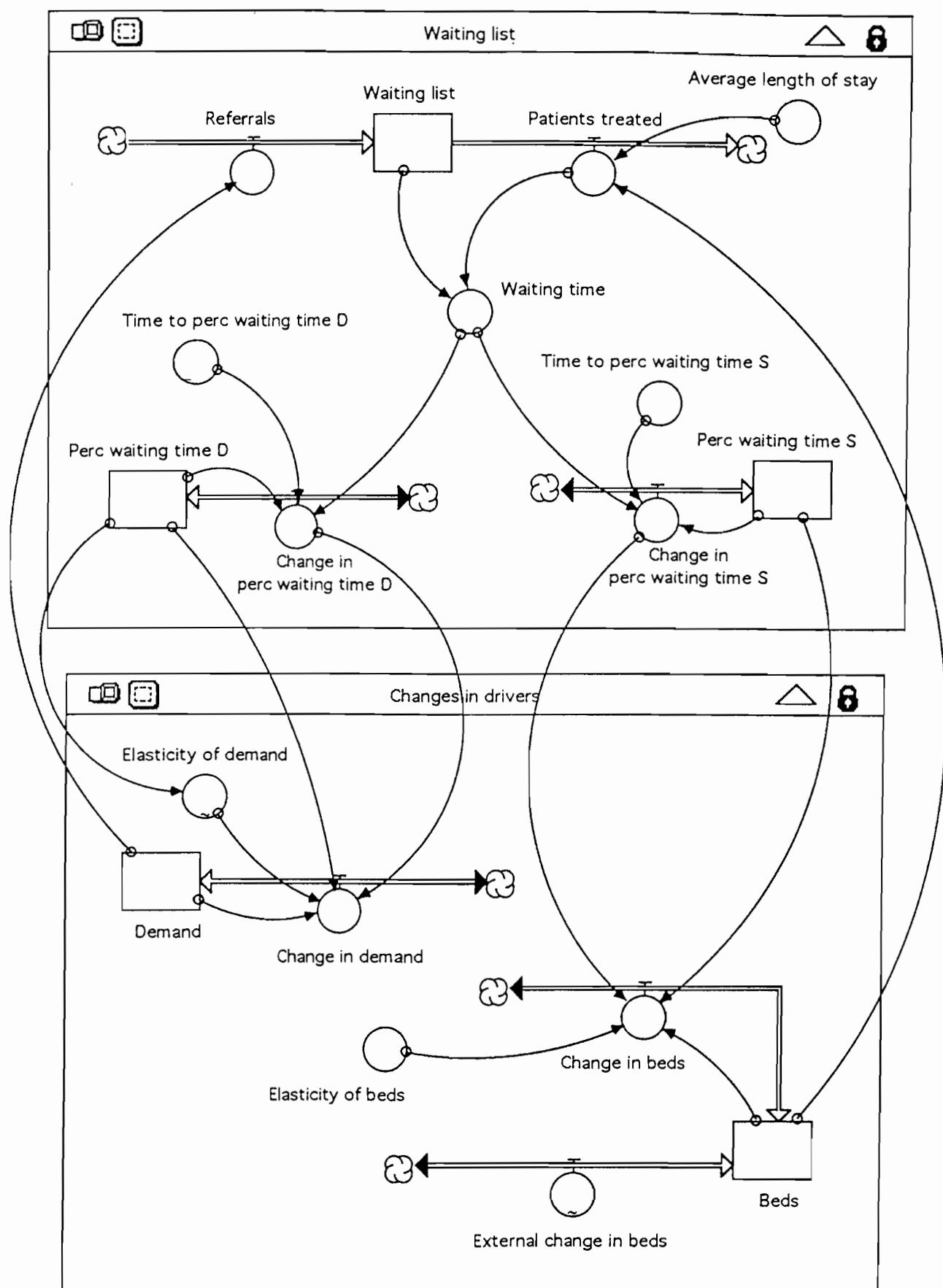


Figure 3 (a). Elasticity of demand with respect to average waiting time as perceived by the demand side

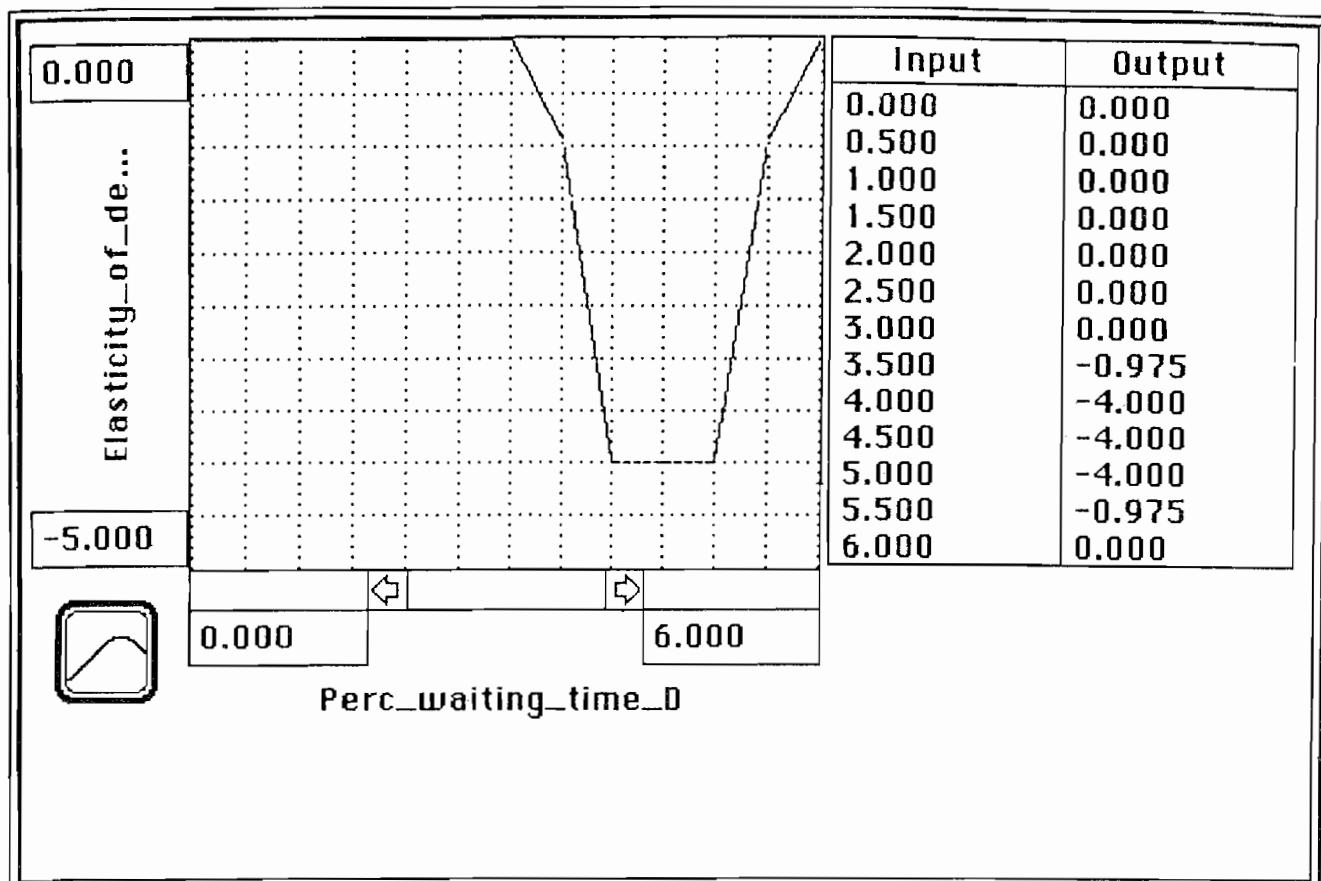


Figure 3(b). Implied demand curve

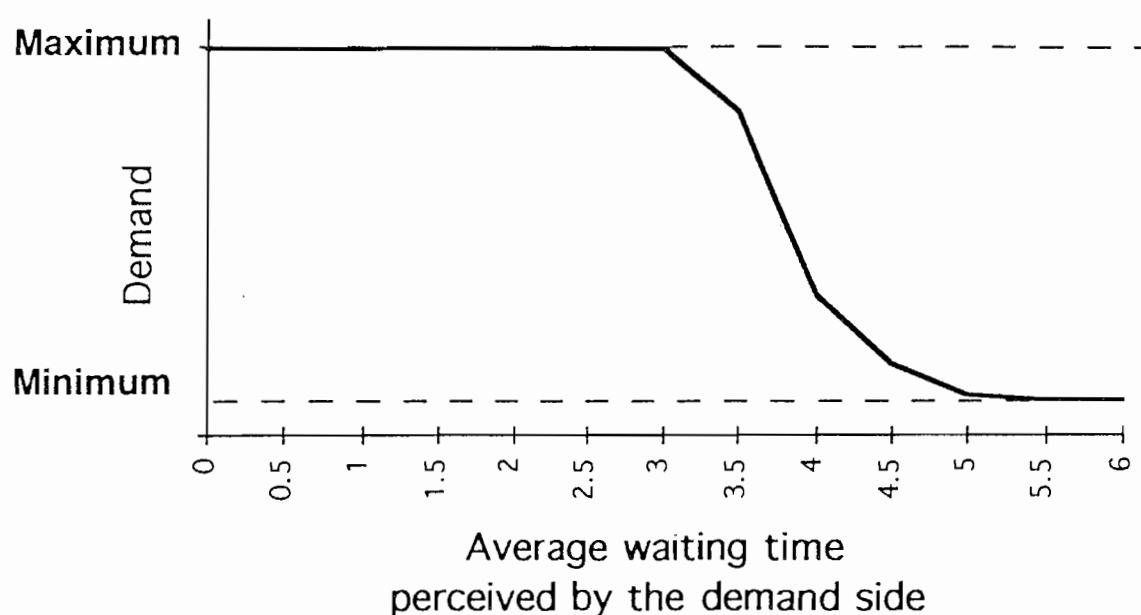


Figure 4(a). Average waiting time for elective surgery

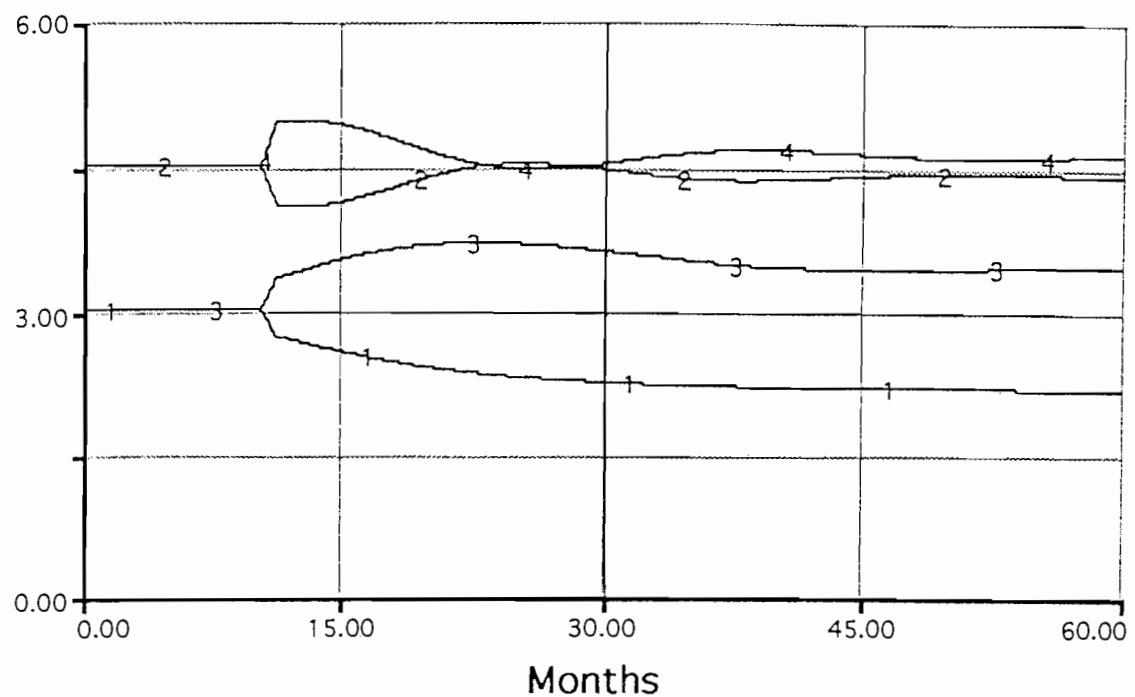
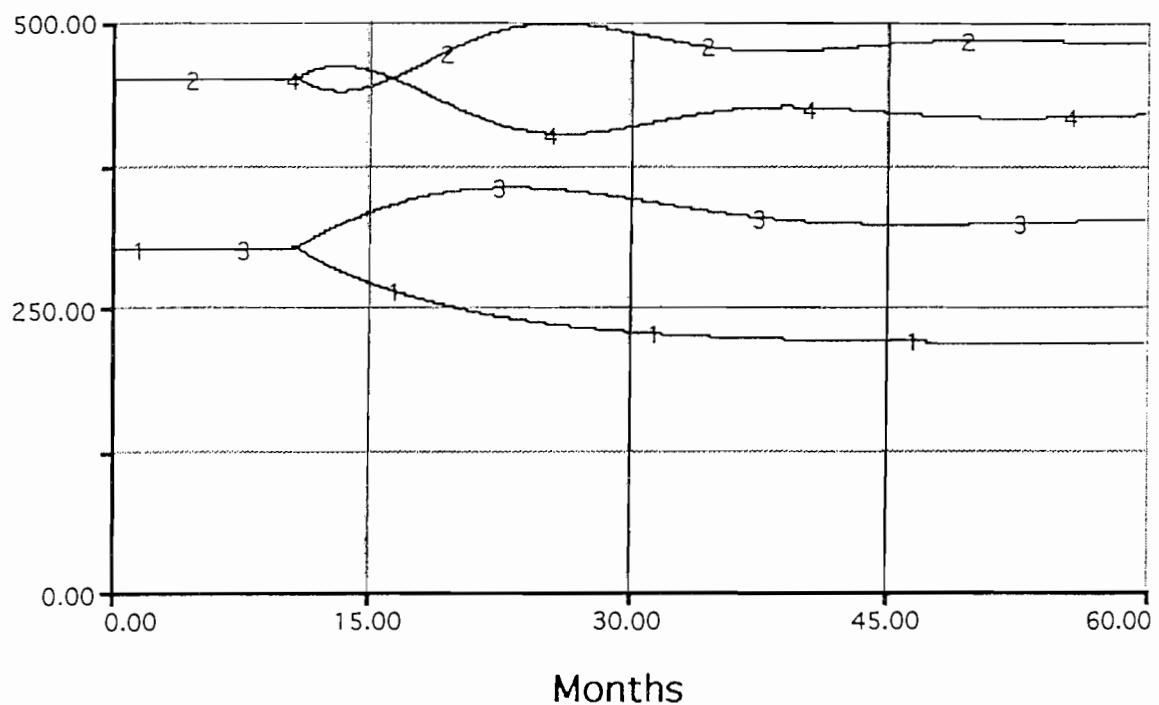


Figure 4(b). Waiting list for elective surgery



Legend

1. Waiting time = 3 months, 10% increase in resources after 10 months
2. Waiting time = 4.5 months, 10% increase in resources after 10 months
3. Waiting time = 3 months, 10% decrease in resources after 10 months
4. Waiting time = 4.5 months, 10% decrease in resources after 10 months

Figure 4(c) Beds for elective surgery

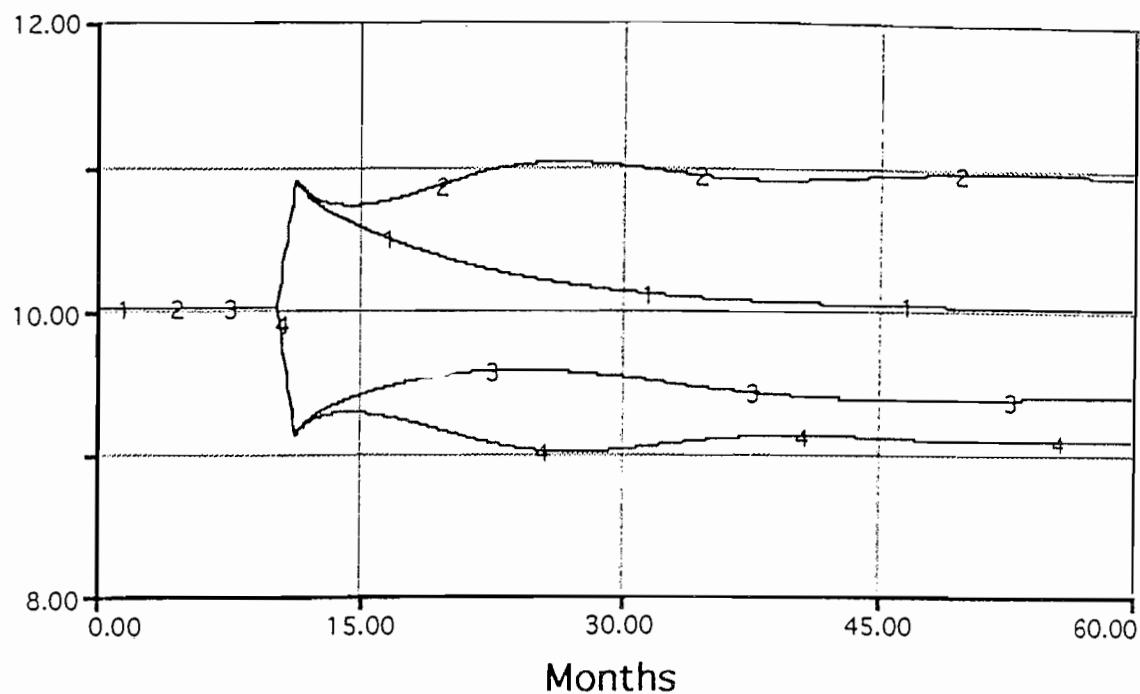
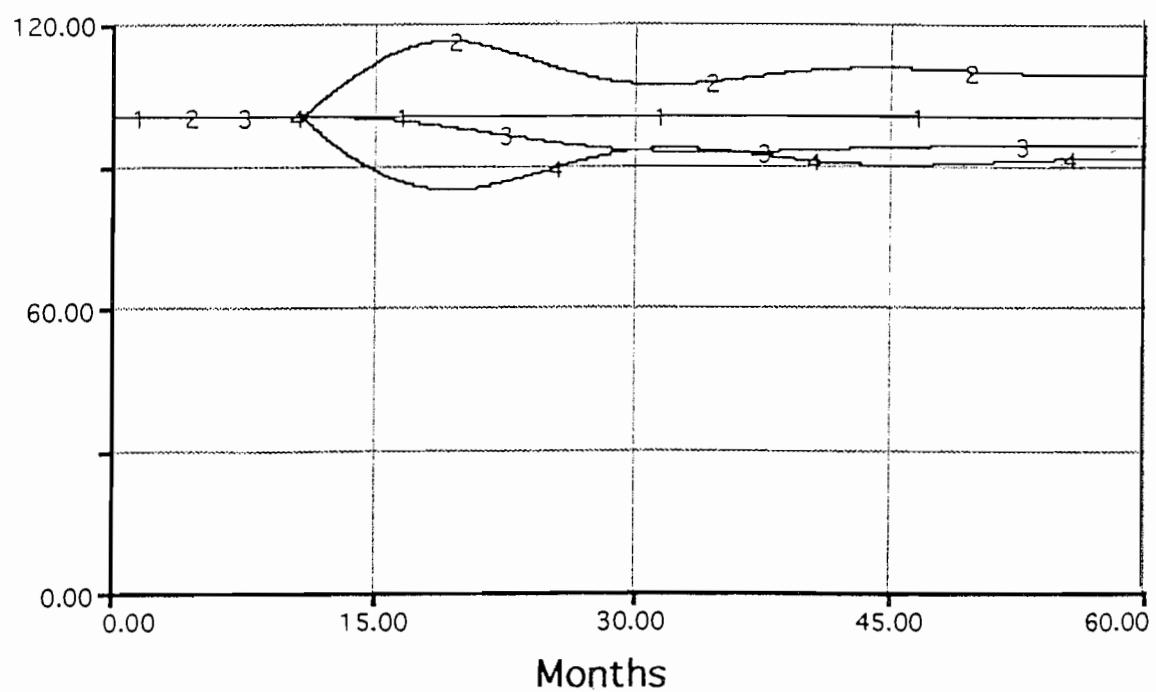


Figure 4(d). Demand for elective surgery (patients per month)



Legend

1. Waiting time = 3 months, 10% increase in resources after 10 months
2. Waiting time = 4.5 months, 10% increase in resources after 10 months
3. Waiting time = 3 months, 10% decrease in resources after 10 months
4. Waiting time = 4.5 months, 10% decrease in resources after 10 months