RM-78-67

THE IIASA HEALTH CARE RESOURCE ALLOCATION SUBMODEL: ESTIMATION OF PARAMETERS

David J. Hughes

December 1978

Preface

The aim of the IIASA Modelling Health Care Systems Task is to build a family of models for the National Health Care System, and to apply them in collaboration with national research centres as an aid to Health Service planners. The modelling work is proceeding along the lines proposed in earlier papers. It involves the construction of linked submodels dealing with population, disease prevalence, resource need, resource allocation, and resource supply.

This paper studies the problems of estimating the parameters of the resource allocation submodel. Earlier procedures are further developed to give methods which have wider application in the planning of health services, and which make direct use of historical allocation data. These procedures are available as computer programs, and three illustrative examples of their use are presented.

Recent related publications of the IIASA Modelling Health Care Systems Task are listed on the back pages of this Memorandum.

Evgenii N. Shigan
Leader
Health Care Systems
Task

November 1978

Abstract

The function of the resource allocation submodel within the IIASA Health Care System model is to simulate how the HCS allocates limited supplies of resources between competing demands. The principal outputs of the submodel are the numbers of patients treated, in different categories, and the modes and quotas of treatment they receive. This paper reviews the data which are available for estimating the parameters of the model, and develops methods which made direct use of historical allocation data. Separate procedures are developed for estimating elasticities, ideal levels of care, and resource costs. These procedures have been realized as computer programs, and their use is illustrated by three examples using hospital data.

ACKNOWLEGEMENT

I am grateful to many colleagues at IIASA and in London for their help in this work. Also I thank the officers of the UK South Western Regional Health Authority for the information used in Example 3 and Rebecca Crow for the typing and artwork.

Contents

The IIASA Health Care Resource Allocation Submodel: Estimation of Parameters

1. INTRODUCTION

The disaggregated resource allocation model DRAM is one of the sub-models of the health care system (HCS) model conceived by Venedictov and Shigan [1], and now being developed by a group of scientists from different countries working at the International Institute for Applied Systems Analysis. Like the other submodels which deal with population, morbidity, resource need and resource supply, DRAM is designed for application by collaborating research centres as an aid to health service planning. Mark 1 [2] and Mark 2 [3] versions of DRAM have already been established, and a comprehensive Mark 3 model has been formulated [4]. In this paper, the methods of parameter estimation which were developed for the earlier versions of DRAM are combined and extended to DRAM Mark 3.

This first section reviews the problems involved in estimating the parameters of a resource allocation model, and motivates the approach developed in the rest of the paper.

1.1 Model Parameters

The purpose of DRAM is to model how the HCS satisfies needs for health care with limited resources. The parameters of the model fall into three groups:

a) the *ideal levels* at which patients would be admitted and receive resources, if there were no constraints on resource availability. These parameters indicate the true "needs" for health care. For example, we might assume that each patient with varicose veins needs, on average, 15 days in-patient hospitalization.

b) the *elasticities* of the actual levels to changes in resource supply. These parameters indicate how the HCS balances need with supply. For example, we expect the elasticity of admission rate to bed availability to be lower for appendicitis patients than for bronchitis patients, because the former condition usually requires faster attention.

c) the *relative costs* of different resources. DRAM uses the marginal unit cost of a bed-day, a doctor-hour, etc., or equivalent parameters, in order to choose between alternative mixes of these resources.

The level of available resources is not regarded as a model parameter but as an experimental variable. DRAM shows how the levels of satisfied demand vary with changes in resource supply.

1.2 Sources of Data

There are more data available to estimate these parameters than there are for many other problems in HCS modelling. We can identify four sources:

a) other models
b) special surveys
c) professional opinion
d) routine statistics

At IIASA, *other models* have been developed for other components of the HCS, and particularly for the estimation of true morbidity from degenerative [5] and infectious [6] diseases. At a later stage in our work, these outputs may be useful for setting the ideal rates at which patients in different categories need treatment. Initially, however, we wish to test and use DRAM independently of other models. Many researchers have performed important and useful *special surveys*. Among many others, Newhouse [7] and Feldstein [8] have estimated both elasticities in hospital care and the costs of acute services, and some of these results were used to calibrate a version of DRAM Mark 1 [2]. Unfortunately, these results may not be relevant in other regions or countries, or at other times. Especially in an international setting it is necessary to avoid reliance on work specific to a specific health system.

The *professional opinions* of doctors and health planners can be useful for setting ideal levels of care. Countries where there is a strong degree of central planning often set normative figures for ideal hospitalization rates and necessary standards of care [9] and these can be used in DRAM. However, these are not available in all countries, and probably no professional should be asked to estimate elasticities, in case he supplies his own rather than those of the HCS. We turn then to *routine statistics*. Most HCSs keep regular records on the use and costs of their services, and on how they have allocated resources in the past. If DRAM is a valid model of this process, then these figures are typical outputs of the model, which we should be able to use for model calibration. This is not to imply that the other sources will never be useful: only that we need to have examined methods for parameters estimation which do not rely on other sources.

The aim of DRAM is to model how the HCS reacts to change. Generally, therefore, DRAM's model parameters must be estimated from data which themselves record change, either in space or time. *Cross-sectional* data from subregions of the region of interest may show the HCS operating at different resource levels. So also may *longitudinal* data collected at different times. In both cases, however, the underlying system may be different for the different data. Subregions are often deliberately defined so as to be predominately urban or predominately rural, and we must consider ways of averaging the results across the region. Data collected at different times are highly likely to be affected by historic trends in medicine or management. Ideally, we should model these trends and incorporate the time-varying parameters in a time-dependent model. More probably, we shall use data from a period during which we can assume time variations to be small. The resulting model will still be good for representing those aspects of resource allocation behaviour which are independent of time trends. A final and obvious problem is that the available data may be incomplete, either because of recording failures or because the data is insufficiently disaggregated.

1.3 Scope of This Paper

Not all of these problems can be overcome simultaneously. However, after the brief *model statement* given in Section 2, Section 3 concentrates on *estimation methods* which

a) are based on routine statistics about current or past allocation behaviour, and

b) recognise that cross-sectional and longitudinal data may reflect inherent parameter variations.

In addition, one of the procedures can be used with incomplete data. Section 4 illustrates the use of these methods on *data* from England and Czechoslovakia, and Section 5 *concludes*.

2. MODEL STATEMENT

This section briefly summarises (from [4]) the version of DRAM Mark 3 for which we desire parameter estimation procedures. There are two *model variables*:

x_{jk} = numbers of individuals in the j-th patient category who receive the k-th mode of treatment (per head of population, per year)

$y_{jk\ell}$ = amounts or quotas of resource type ℓ received by each individual in the j-th patient category treated in the k-th mode.

The model chooses x and y* so as to maximise a *utility function*

$$U(x,y) = \sum_j \sum_k g_{jk}(x_{jk}) + \sum_j \sum_k \sum_\ell x_{jk} h_{jk\ell}(y_{jk\ell}) \qquad (1)$$

where

---

*In the sequel, we use x,y to denote $\{x_{jk}, j=1,2,...J, k=1,2,...K\}$, $\{y_{jk\ell}, j=1,2,...J, k=1,2,...K, \ell=1,2,...L\}$ respectively, with a like notation for similarly subscripted variables.

$$g_{jk}(x) = -\frac{X_{jk}\sum_{\ell} C_\ell Y_{jk\ell}}{\alpha_j}\left(\frac{x}{X_{jk}}\right)^{-\alpha_j} \quad (2)$$

$$h_{jk\ell}(y) = \frac{C_\ell Y_{jk\ell}}{\beta_{jk\ell}}\left\{1 - \left\{\frac{y}{Y_{jk\ell}}\right\}^{-\beta_{jk\ell}}\right\} \quad (3)$$

within the *resource constraint*

$$\sum_j \sum_k x_{jk} y_{jk\ell} = R_\ell \qquad \forall\ \ell\ . \quad (4)$$

The availabilities R_ℓ of resource type ℓ are assumed to be given exogenously. The unit costs C_ℓ, the model elasticities α_j, $\beta_{jk\ell}$, and the ideal levels $X_{jk}, Y_{jk\ell}$, which are upper bounds on the model variables, are *parameters* which have to be estimated. Figure 1 shows how the model is used in solution mode.

In the normal way we adjoin the L constraint equations (4) to the utility function which is to be maximised (1) by means of

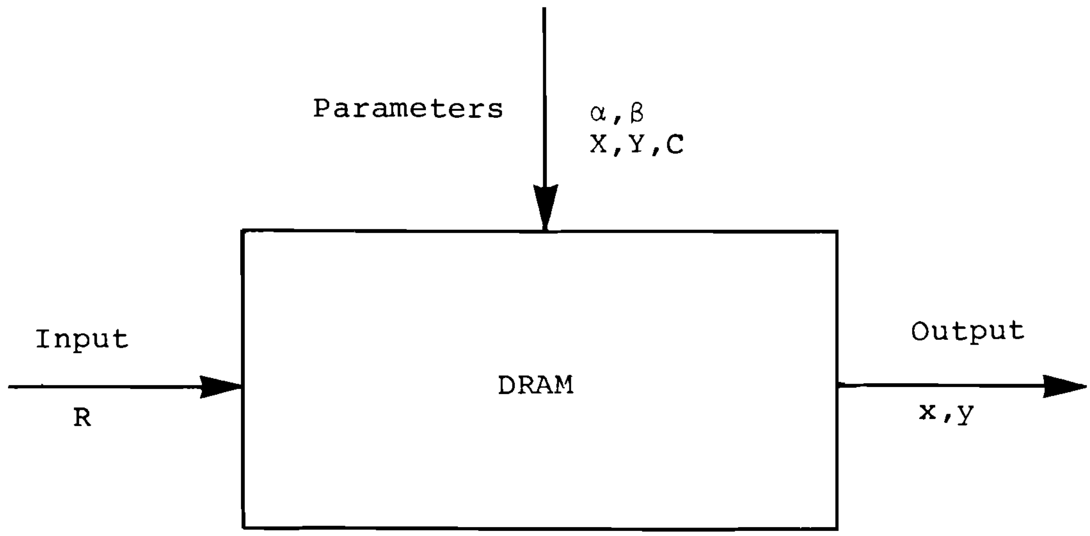

Figure 1. DRAM solves x,y for different values of R.

L arbitrary multipliers λ_ℓ. For convenience we scale these by the cost of each resource type C_ℓ

$$H(x,y,\lambda) = \sum_j \sum_k g_{jk}(x_{jk}) + \sum_j \sum_k \sum_\ell x_{jk} h_{jk\ell}(y_{jk\ell})$$

$$+ \sum_\ell C_\ell \lambda_\ell (R_\ell - \sum_j \sum_k x_{jk} y_{jk\ell}) \quad . \tag{5}$$

In order to find the values of x and y which maximise H, we must solve the JK(L + 1) + L equations

$$\frac{\partial H}{\partial x_{jk}} = \frac{\partial H}{\partial y_{jk\ell}} = \frac{\partial H}{\partial \lambda_\ell} = 0 \qquad \forall\ j,k,\ell \tag{6,7,8}$$

for the JK(L + 1) + L unknowns: x,y, and λ. First,

$$\frac{\partial H}{\partial y_{jk\ell}} = x_{jk} h'_{jk\ell}(y_{jk\ell}) - C_\ell \lambda_\ell x_{jk} = 0$$

leads to

$$y_{jk\ell} = Y_{jk\ell}(\lambda_\ell)^{\frac{-1}{\beta_{jk\ell}+1}} \quad . \tag{9}$$

Secondly

$$\frac{\partial H}{\partial x_{jk}} = g'_{jk}(x_{jk}) + \sum_\ell h_{jk\ell}(y_{jk\ell}) - \sum_\ell C_\ell \lambda_\ell y_{jk\ell} = 0$$

leads to

$$x_{jk} = X_{jk}(\mu_{jk})^{\frac{-1}{\alpha_j+1}} \tag{10}$$

where μ_{jk} is a weighted sum

$$\mu_{jk} = \frac{\sum_{\ell} C_{\ell} Y_{jk\ell} \nu_{jk\ell}}{\sum_{\ell} C_{\ell} Y_{jk\ell}} \qquad (11)$$

of the terms

$$\nu_{jk\ell} = \left((\beta_{jk\ell} + 1) \lambda_{\ell}^{\frac{\beta_{jk\ell}}{\beta_{jk\ell}+1}} - 1 \right) / \beta_{jk\ell} \quad . \qquad (12)$$

Substituting the results of equations (9) and (10) into the constraint equation (4) gives

$$f_{\ell}(\lambda_1, \lambda_2, \ldots \lambda_L) = f_{\ell}(\lambda) = 0 \qquad \forall\ \ell \qquad (13)$$

where

$$f_{\ell}(\lambda) = -R_{\ell} + \sum_{j} \sum_{k} X_{jk} Y_{jk\ell} (\lambda_{\ell})^{\frac{-1}{\beta_{jk\ell+1}}} (\mu_{jk})^{\frac{-1}{\alpha_j+1}} \qquad (14)$$

which must be solved for λ by a numerical technique such as the multi-dimensional extension of the Newton-Raphson method. In this method, an approximate solution $\hat{\lambda}$ yields an improved solution λ according to

$$\lambda_{\ell} = \hat{\lambda}_{\ell} - \sum_{m} \bar{D}_{\ell m} f_m(\lambda) \qquad \forall\ \ell \qquad (15)$$

where $\bar{D}_{\ell m}$ is the ℓm-th element of the matrix

$$\underline{\bar{D}} = \underline{D}^{-1} = \left\{ \frac{\partial f_{\ell}(\lambda)}{\partial \lambda_m} \right\}^{-1} \qquad (16)$$

which is the inverse of the matrix $\underline{D}$ of partial derivatives of $f(\lambda)$. These partial derivatives are

$$\frac{\partial f_\ell(\lambda)}{\partial \lambda_m} = -\sum_j \sum_k T_{jk\ell m} X_{jk} Y_{jk\ell} (\lambda_\ell)^{\frac{-1}{\beta_{jk\ell}+1}-1} (\mu_{jk})^{\frac{-1}{\alpha_j+1}-1} \qquad (17)$$

where

$$T_{jk\ell m} = \left(\frac{\mu_{jk}}{\beta_{jk\ell}+1} \cdot \frac{\partial \lambda_\ell}{\partial \lambda_m} \right) + \left(\frac{\lambda_\ell}{\alpha_j+1} \cdot \frac{\partial \mu_{jk}}{\partial \lambda_m} \right)$$

$$\frac{\partial \lambda_\ell}{\partial \lambda_\ell} = 1 \quad , \qquad \frac{\partial \lambda_\ell}{\partial \lambda_m} = 0 \qquad \text{for } m \neq \ell$$

and

$$\frac{\partial \mu_{jk}}{\partial \lambda_m} = \frac{C_m Y_{jkm}}{\sum_m C_m Y_{jkm}} (\lambda_m)^{\frac{-1}{\beta_{jkm}+1}} \; . \qquad (18)$$

This approach is equivalent to solving the dual of the original optimisation problem. Appendix 1 explains the relation.

3 ESTIMATION METHODS

This section describes various methods for estimating the model parameters α, β, X, Y, C of DRAM. We give three groups of methods for estimating α, β, X, Y when the unit costs C are known. They are:

a) A *combined calibration and validation* approach which chooses all the model parameters so as to minimise the total discrepancy between the model and all the available data. This method, although comprehensive, is probably too complicated for practical use.

b) Methods for estimating α, β *when X, Y are known or given exogenously*. We can choose α, β simply so that the model reproduces the current allocation of resources, or on the basis of information about the elasticities of output to supply.

c) Methods for estimating X,Y *when* α,β *are known or given exogenously*. Again we can use just the current allocation of resources, or more detailed cross-sectional or longitudinal data. In the latter case, however, we must consider that the model parameters may change in space or time.

Given sufficient data, the latter two methods may be combined in an iterative approach. Finally in this section, we look separately at methods for estimating the resource costs C. The problems here are mainly definitional.

3.1 Combined Calibration and Validation

First we describe what might be the ideal method for estimating model parameters, if it could be implemented. This would be a procedure which takes a large amount of data on comparable historic resource allocations and which derives the best parameter estimates, together with measures of the goodness of fit between the data and the model hypotheses. The structure of DRAM is such that there is a natural way to formulate this task in mathematical terms, although it is less easy to see how to implement it.

To illustrate the approach, consider a DRAM with one category, one mode, and one resource ($J = K = L = 1$). Figure 2 shows the locus OA of possible model solution on the xy plane, for three different parameter sets. The solution for a given resource level R is given by the intersection of the locus with the constant resource hyperbola $xy = R$. On each line we have the model outputs (circles) for some resource levels, and nearby on the same hyperbolae are the observed outputs (crosses). We see that

a) along O1A an appropriate choice of the model parameters α,β,X,Y has aligned the circles and the crosses,

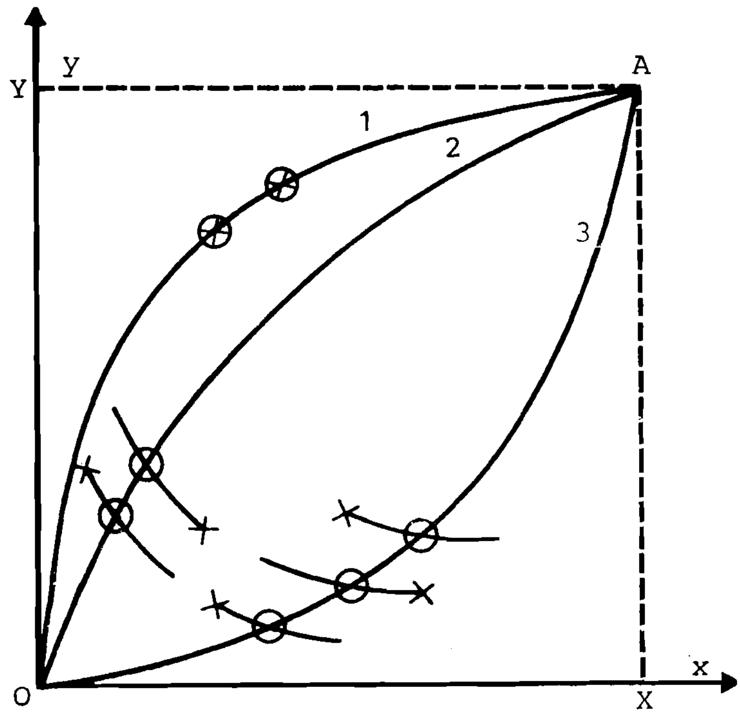

Figure 2. Model outputs (o) and observed outputs (x) for three different parameter sets.

b) along O2A, the circles and the crosses do *not* coincide, because the two observed outputs are such that they *cannot* be reproduced by *any* choice of model parameters (Appendix 2 summarizes some results on admissible solutions for DRAM),

c) along O3A, the circles and the crosses do *not* coincide, because DRAM has insufficient parameters to fit the solution line to three arbitrary points. With any number of data points more than two, the estimation problem is generally *overspecified*.

The common feature of b) and c) is that, for any choice of parameters, the utility achieved by the model for a particular value of R will always exceed the utility of the observed values of x,y. This difference in utility is a measure of the disagreement between the data and the hypothesis that the data maximises a utility function like equation (1). If the model parameters α, β, X,Y can be chosen so that this difference is acceptably small, then the model can be calibrated and validated simultaneously.

By defining the N observation pairs x(i),y(i), i = 1,2,...N, we can write the utility of each output pair (x(i),y(i)) as

$$U_o(i) = \frac{CXY}{\alpha}\left(\frac{x(i)}{X}\right)^{-\alpha} + \frac{CXY}{\beta}\left(\frac{x(i)}{X}\right)\left(1 - \left(\frac{y(i)}{Y}\right)^{-\beta}\right) \quad .$$

The utility of the corresponding model outputs is given by substituting equations (9),(10) into (1)

$$U_m(i) = \frac{CXY}{\alpha}\left(\mu(i)\right)^{\frac{\alpha}{\alpha+1}} + \frac{CXY}{\beta}\left(\mu(i)\right)^{\frac{-1}{\alpha+1}}\left(1 - \left(\lambda(i)\right)^{\frac{\beta}{\beta+1}}\right)$$

where $\mu(i)$ is related to $\lambda(i)$ by equations (11), (12) and $\lambda(i)$ satisfies

$$f\left(\lambda(i)\right) = - x(i)y(i) + XY\left(\lambda(i)\right)^{\frac{-1}{\beta+1}}\left(\mu(i)\right)^{\frac{-1}{\alpha+1}} = 0 \quad . \qquad (19)$$

The utility difference associated with a single observation is therefore

$$\Delta U(i) = U_m(i) - U_o(i)$$

$$= CXY\left\{\begin{array}{l} \frac{-1}{\alpha}\left[\left(\mu(i)\right)^{\frac{\alpha}{\alpha+1}} - \left(\frac{x(i)}{X}\right)^{-\alpha}\right] \\ + \frac{1}{\beta}\left[\begin{array}{l}\left(\mu(i)\right)^{\frac{-1}{\alpha+1}}\left(1 - \left(\lambda(i)\right)^{\frac{\beta}{\beta+1}}\right) \\ -\left(\frac{x(i)}{X}\right)\left(1 - \left(\frac{y(i)}{Y}\right)^{-\beta}\right)\end{array}\right]\end{array}\right\} \quad . \qquad (20)$$

Because the model outputs are those which maximise $U_m(i)$, we always have $\Delta U(i) \geq 0$, and the disagreement between N observed data points and a particular set of parameters is measured by

$$\Delta U_N = \sum_{i=1}^{N} \Delta U(i) \geq 0 \quad . \tag{21}$$

We can now define the task. It is to choose α, β, X, Y so as to minimise ΔU_N. A minimised value of ΔU_N which is small will indicate a good fit between data and model, and vice versa. Although we have restricted ourselves to the case when $J = K = L = 1$, the extension to many categories, modes, and resources is straightforward. The problem lies in the fact that this *unconstrained* minimisation problem is not suitably convex. No amount of data gives information about the Lagrange multipliers for a given resource level, or about the relative sizes of the multipliers for different resource levels. Just as in the estimation procedure used for DRAM Mark 2 [3], constraints must be applied upon the scale and shape of admissable solutions in order to avoid obtaining pathological results such as $\alpha = \beta = 0$. The resulting *constrained* minimisation problem is rather intractable although perhaps not impossible for small models. Appendix 3 summarizes some proved results and some of the outstanding difficulties in this approach.

3.2 Estimation of α, β

Because what might be the ideal estimation method is generally impractical, we must consider other approaches. To simplify the exposition, we assume first that the ideal admission rates X are available from morbidity prediction models, analyses of hospitalization trends, or are otherwise known exogenously. We similarly assume that the ideal resource quotas Y can be specified exogenously by medical professionals. This leaves only the elasticities α, β to be estimated.

Sufficient information to estimate α, β is given by the current allocation of resources in the region under study. If the current allocation pattern is described by x and y, equations (9) and (10) may be rearranged as

$$\alpha_j = \left(\ell n(\mu_{jk})/\ell n\left(\frac{X_{jk}}{x_{jk}}\right)\right) - 1 \tag{22}$$

$$\beta_{jk\ell} = \left(\ell n(\lambda_\ell)/\ell n\left(\frac{Y_{jk\ell}}{y_{jk_\ell}}\right)\right) - 1 \tag{23}$$

which are expressions for α and β. Two small problems must be overcome. First, we know from [3] that λ must be determined externally. We know also, however, that α and β are always positive. This implies then that

$$\lambda_\ell > \tilde{\lambda}_\ell = \max_{j,k} \left\{ \left(\frac{X_{jk}}{x_{jk}}\right), \left(\frac{Y_{jk\ell}}{y_{jk\ell}}\right) \right\} \qquad \forall\ \ell \tag{24}$$

and we can conveniently define λ_ℓ as some (small) multiple $\phi_\ell > 1$ of the minimum value $\tilde{\lambda}_\ell$

$$\lambda_\ell = \phi_\ell \tilde{\lambda}_\ell \qquad \forall\ \ell\ . \tag{25}$$

The second problem is that equation (22) gives K values for each α_j. Generally these will be different values, but we can overcome this by aggregating the data across modes, and by using equations (22), (23) with k = 1.

By these means, we may estimate values for the parameters α, β. The model so calibrated will not exactly reproduce the current allocation of resources unless the latter is one of the admissible solutions of DRAM defined in Appendix 2. However, it will reproduce the actual quotas $y_{jk\ell}$, and the actual numbers of patients treated in each category $(x_{j1} + x_{j2} + \ldots + x_{jK})$. Whether the estimated elasticities are useful for forward prediction will depend upon whether the current allocation pattern is representative of the HCS's usual behaviour. The procedure described above only finds values for α, β which are consistent with this assumption and with the values assumed for X,Y.

A more sophisticated approach is to make use of more data by estimating *empirical elasticities*. These can then be used to derive the model elasticities α, β. Appropriate empirical elasticities for DRAM Mark 3 are

$\gamma_{jk\ell}$ = the elasticity of the admission rate x_{jk} to changes in the resource level R_ℓ,

$\eta_{jkm\ell}$ = the elasticity of the resource quota y_{jkm} to changes in the resource level R_ℓ.

These quantities are readily expressed in terms of α, β. For example, $\gamma_{jk\ell}$ is

$$\gamma_{jk\ell} = \frac{\partial \ell n\, x_{jk}}{\partial \ell n\, R_\ell} = \frac{\partial \ell n\, x_{jk}}{\partial \mu_{jk}} \cdot \frac{\partial \mu_{jk}}{\partial R_\ell} \cdot R_\ell$$

and using (10) to give an expression for $\partial \ell n\, x_{jk}/\partial \mu_{jk}$ yields

$$\gamma_{jk\ell} = \frac{-R_\ell}{(\alpha_j + 1)\mu_{jk}} \cdot \frac{\partial \mu_{jk}}{\partial R_\ell} \quad . \tag{26}$$

Similarly

$$\eta_{jkm\ell} = \frac{-R_\ell}{(\beta_{jkm} + 1)\lambda_m} \cdot \frac{\partial \lambda_m}{\partial R_\ell} \tag{27}$$

where

$$\frac{\partial \mu_{jk}}{\partial R_\ell} = \sum_m \frac{\partial \mu_{jk}}{\partial \lambda_m} \cdot \frac{\partial \lambda_m}{\partial R_\ell} \quad . \tag{28}$$

As in [3], we may show that the derivatives

$$\frac{\partial R_\ell}{\partial \lambda_m} = \frac{\partial f_\ell(\lambda)}{\partial \lambda_m}$$

are related to those of equation (17). However, although it is straightforward to express γ,η in terms of α,β, it is impossible to express α,β explicitly in terms of γ,η. This is because the various partial derivatives in these formulae depend upon α,β in such a way that they cannot be inverted. Instead, we write equations (26), (27) as

$$\alpha_j = \frac{A_{jk\ell}}{\gamma_{jk\ell}} - 1 \tag{29}$$

$$\beta_{jkm} = \frac{B_{m\ell}}{\eta_{jkm\ell}} - 1 \tag{30}$$

where

$$A_{jk\ell} = \frac{-R_\ell}{\mu_{jk}} \sum_m \left(\frac{\partial \mu_{jk}}{\partial \lambda_m} \right) \bar{D}_{m\ell} \tag{31}$$

$$B_{m\ell} = \frac{-R_\ell}{\lambda_m} \bar{D}_{m\ell} \tag{32}$$

which, if γ,η have been determined in some other study (such as [8]), and if λ is set by an equation like (25), may be solved iteratively for α,β.

There are nevertheless some problems in this approach. First, it is necessary to ensure that the empirical elasticities are consistent with the choice of λ, otherwise the procedure may not converge. Secondly, there are more empirical elasticities γ,η than there are model elasticities α,β. Therefore, unless some of the empirical elasticities are ignored, the model parameters will be overspecified. Thirdly, the empirical elasticities γ,η are not directly measurable and are usually the result of some prior data analysis. For example, estimates $\hat{\gamma},\hat{\eta}$ can be found by assuming that some N known data points $x_{jk}(i)$, $y_{jk\ell}(i)$, $R_\ell(i)$, $i = 1,\ldots N$, satisfy the linear models

$$\ell n\, x_{jk}(i) = a^x_{jk} + \sum_\ell \gamma_{jk\ell}\,\ell n\left(R_\ell(i)\right) + \varepsilon^x_{jk}(i) \tag{33}$$

$$\ell n\, y_{jkm}(i) = a^y_{jkm} + \sum_\ell \eta_{jkm\ell}\,\ell n\left(R_\ell(i)\right) + \varepsilon^y_{jkm}(i) \tag{34}$$

in which a^x, a^y are unknown constants, and in which $\varepsilon^x, \varepsilon^y$ are random, uncorrelated, error terms with zero means. Equations (33), (34) are only approximately true, because they imply that γ, η do not change as the resource level changes, and equations (26), (27) contradict this. Nevertheless, if we eliminate γ, η by combining equations (29), (30), (33), (34) to give

$$\ell n\left(x_{jk}(i)\right) = a^x_{jk} + \left(\frac{1}{\alpha_j + 1}\right) \sum_\ell A_{jk\ell}\,\ell n\left(R_\ell(i)\right) + \varepsilon^x_{jk}(i)$$

for each j, for all k,i, (35)

and

$$\ell n\left(y_{jkm}(i)\right) = a^y_{jk\ell} + \left(\frac{1}{\beta_{jkm} + 1}\right) \sum_\ell B_{m\ell}\,\ell n\left(R_\ell(i)\right) + \varepsilon^y_{jk\ell}(i)$$

for each j,k,m, for all i, (36)

which are linear equation of the form

$$\zeta(i) = a + b\chi(i) + \varepsilon(i) \tag{37}$$

we can use the following iterative scheme in order to estimate α and β.

a) Fix λ arbitrarily for some resource level R, perhaps by using equation (25) on one of the data points.

b) Assume some initial estimates of α, β (e.g., unity).

c) Derive μ from equations (11), (12), $\frac{\partial f_\ell}{\partial \lambda_m}$ from equation (17), and A,B from equations (31), (32).

d) Find the best least-squares estimators of $(\alpha_j + 1)^{-1}$, $(\beta_{jkm} + 1)^{-1}$ in equations (35),(36).

e) Hence, estimate α, β and repeat from c) until convergence.

This procedure (also depicted in Figure 3) is likely to be more lengthy than the procedures used in earlier versions of DRAM, because it incorporates the regression estimation of empirical elasticities, which was previously performed separately. On the other hand, it has the advantage that more of the original data can be used directly. If a full data set

$$\{x_{jk}(i), y_{jk\ell}(i), R_\ell(i);\ i = 1,\ldots N,\ j = 1,\ldots J,\ k = 1,\ldots K,\ \ell = 1,\ldots L\}$$

is available, KN equations are available to estimate each α_j, and probably not all of the $x_{jk}(i)$ need be known. Fewer equations (just N) are available to estimate each $\beta_{jk\ell}$, and it may be necessary to introduce some further simplifying assumptions such as

$$\beta_{j_1 k\ell} = \beta_{j_2 k\ell} \qquad \forall\ j_1, j_2 \in \{1,\ldots J\} \tag{38}$$

in order to obtain reliable estimates. A second advantage of this procedure is that it is not necessary to modify any of the input data to make them consistent with the model. A third advantage is that the parameter estimated in each regression has an estimated standard error associated with it. These errors provide a measure of the reliability of α, β.

Perhaps the main assumption in the above analysis is that the underlying elasticities are constant across the set of data points. Because there is little information about how elasticities are likely to vary in time or space, we have not attempted to model this variation here. But Appendix 4 shows that in a certain sense, the procedure described above gives *unbiassed* estimates of the underlying "mean" parameters. This is a reassuring result, and

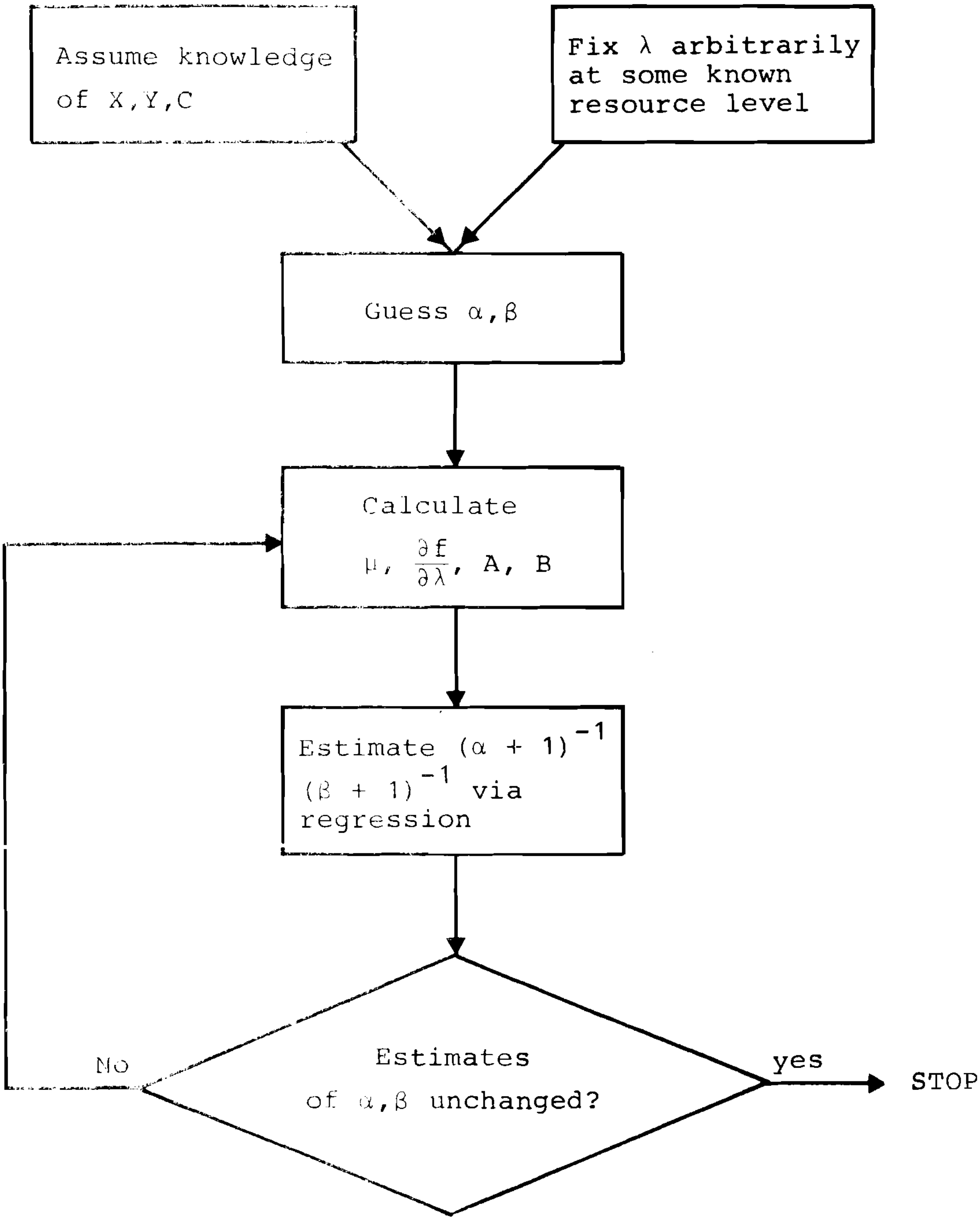

Figure 3. Estimation of elasticities.

the estimates can be further tested to see if the model so calibrated can reproduce the current allocation of resources; data which is not explicitly used for estimation.

3.3 Estimation of X,Y

We turn now to the estimation of the ideal admission rates X and the ideal resource quotas Y, assuming for the moment that the model elasticities α,β are known.

Sufficient information to estimate X,Y is given by the current allocation of resources in the region under study. If the current allocation pattern is described by x and y, equations (9) and (10) may be rearranged as

$$X_{jk} = x_{jk}(\mu_{jk})^{\frac{1}{\alpha_j+1}} \tag{39}$$

$$Y_{jk\ell} = y_{jk\ell}(\lambda_{jk\ell})^{\frac{1}{\beta_{jk\ell}+1}} \tag{40}$$

which are expressions for X and Y. We have a single equation for each unknown parameter, but we must still determine λ by some external criterion. If we assume that we can define the resources needed to satisfy the ideal levels $X_{jk}, Y_{jk\ell}$ as some multiple Θ_ℓ of the resources used currently

$$\sum_j \sum_k X_{jk}Y_{jk\ell} = \Theta_\ell \sum_j \sum_k x_{jk}y_{jk\ell} \qquad \forall\ \ell \tag{41}$$

then (9),(10) can be substituted into (41) to give

$$\tilde{f}_\ell(\lambda) = 0 \qquad \forall\ \ell \tag{42}$$

where

$$\tilde{f}_\ell(\lambda) = -\Theta_\ell \sum_j \sum_k x_{jk}y_{jk\ell} + \sum_j \sum_k x_{jk}y_{jk\ell}(\lambda_\ell)^{\frac{1}{\beta_{jk}+1}}(\mu_{jk})^{\frac{1}{\alpha_j+1}} \tag{43}$$

and where equation (42) must be solved for λ. The equations in $\tilde{f}$ are very similar to equations (13) in f, and provided that $\theta_\ell > 1 \ \forall \ \ell$, and that all the terms except λ are known, they may be solved in the same way to give λ. Unfortunately, not all the terms are known. In particular, μ_{jk} is a weighted average involving the terms $Y_{jk\ell}$, which are as yet unknown. It is therefore necessary to iterate between solving equation (42) for λ, and equations (39), (40) for X,Y.

This approach, like the similar approach described for α,β, suffers from the disadvantage that it only finds values of X,Y which are consistent with the current allocation pattern and the assumed values for α,β. More useful is to estimate X,Y from other data and then to use the current allocation as a test of the model's validity. Other suitable data include cross-sectional and longitudinal data, and given N data points from such sources, we can use equations (39,40) to find N estimates of X,Y. The problem remains of how to combine these estimates.

Estimates $X_{jk}(i), Y_{jk\ell}(i)$ derived for subregions i = 1,...N may be combined rather easily. If the population of the ith subregion is P(i), then

$X_{jk}(i)P(i)$	is the number of individuals in the j-th category in the k-th mode of treatment who need treatment in subregion i (per year) and
$X_{jk}(i)Y_{jk\ell}(i)P(i)$	is the number of resources ℓ needed to treat these individuals (per year).

These quantities may be summed across the region, and the corresponding *regional* estimates of X and Y are

$$\bar{X}_{jk} = \sum_i X_{jk}(i)P(i) \Big/ \sum_i P(i) \qquad \forall \ j,k \tag{44}$$

$$\bar{Y}_{jk\ell} = \sum_i X_{jk}(i)Y_{jk\ell}(i)P(i) \Big/ \sum_i X_{jk}(i)P(i) \qquad \forall \ j,k,\ell \tag{45}$$

This approach (also depicted in Figure 4) is interesting because we do not need to assume that X and Y are constant across the region. The subregional variations are averaged by summing the ideal demands across the region.

Estimates $X_{jk}(i), Y_{jk\ell}(i)$ derived at different times $i = 1,...N$ are more difficult to combine. Ideal resource quotas $Y_{jk\ell}$ are probably decreasing with time, and an exponential curve could be fitted to a long sequence of points. The ideal numbers of patients needing treatment per head of population, $Z_j = \sum_k X_{jk}, \forall j$, will change because of changes in the age structure and in the morbidity rates. The former can be corrected for, and the latter can probably be assumed to be constant. Most difficult to model are the changes in doctors' preferences between modes. These are reflected in the individual values of X_{jk}, which could if necessary be regarded as experimental variables.

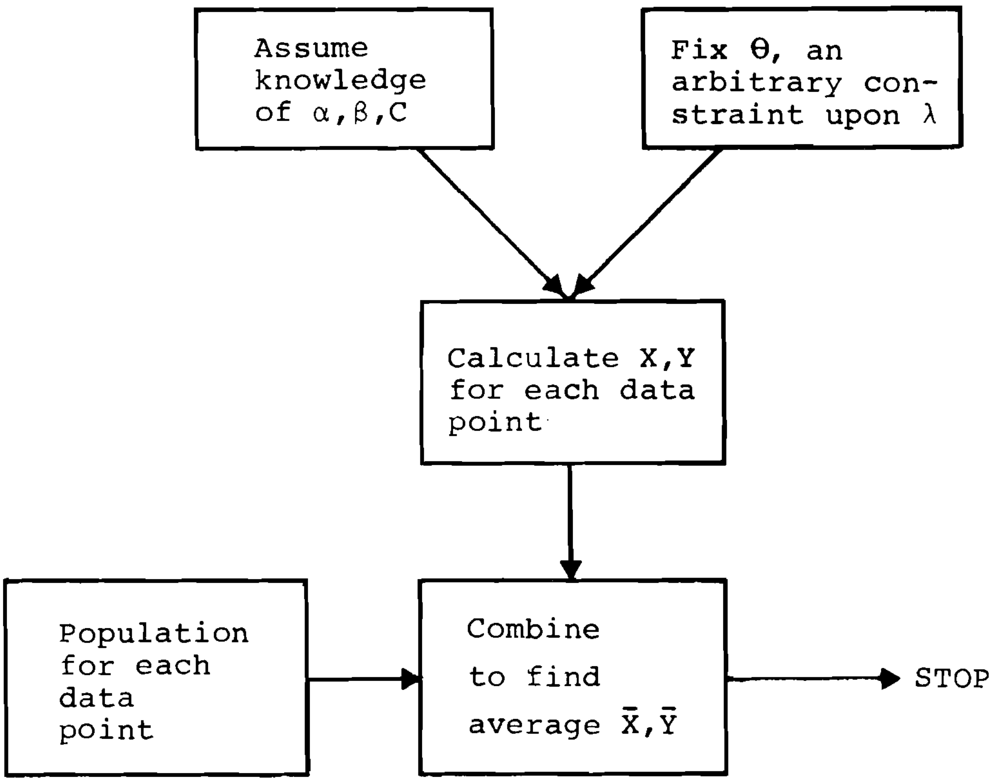

Figure 4. Estimation of ideal levels.

3.4 Estimation of α,β and X,Y

In the most general case, neither of the parameter pairs α,β or X,Y is known, and we require estimates of both. In this circumstance, the two procedures described above may be used together in the following iterative scheme.

a) With some arbitrary initial estimates of X,Y, use the methods of Section 3.2 to estimate α,β.

b) With these estimates of α,β, use the methods of Section 3.3 to estimate X,Y.

c) Repeat from a) until convergence.

The limit to such analysis is set by the amount of data available. The danger of attempting to estimate too many dependencies in time and space is obvious. Less obvious is the danger of using the same data twice to estimate α,β and then X,Y. When neither of these pairs is given exogenously, the same data cannot be used to estimate both pairs of parameters.

All the parameter estimation procedures so far described involve additional constraint variables such as ϕ and Θ which must be fixed arbitrarily. Furthermore the estimates of α,β,X,Y depend strongly upon the choice of ϕ and Θ. Fortunately, however, this is not a problem. Although different values of ϕ,Θ lead to different values for α,β,X,Y, each set of parameter values will reproduce with similar accuracy the data points used for estimation. Provided that predictive runs of the model do not involve resource levels very dfferent from those used in estimation, the results will be relatively insensitive to ϕ,Θ.

3.5 Estimation of C

Finally in this section, we discuss how to estimate the unit resource costs C needed in the model. These parameters are defined rather carefully. Specifically,

C_ℓ = the marginal cost of using one more resource of type ℓ, when all needs for health care are met.

Furthermore, these costs are not money costs but opportunity costs which reflect the benefit in some alternative foregone by buying the extra resource. How then can they be estimated? Often, we have financial data which we can use directly, but when these are unavailable or inappropriate, equivalent model parameters can be inferred from other information.

Two assumptions will enable us to estimate the costs C from financial data, when these are available. The first assumption is that in long-term planning, opportunity costs are approximately measured by money costs. Given sufficient time, every option is an alternative, and all resources are convertible. The second assumption is that marginal costs are approximately measured by average costs. The cost function of an individual hospital or medical school is certainly non-linear, with marginal costs being generally less than average costs. But when many such hospitals or medical schools are operating in a single region, the aggregate cost function may be approximately linear as shown in Figure 5. In these circumstances, the average costs recorded in historic accounts will approximate to the marginal costs at some hypothetical resource level.

However, not all countries compare alternative plans in terms of financial affordability. In the USSR, for example, planning seeks mainly to reconcile the real outputs between producers while satisfying certain aims such as full employment, constant growth, etc. For application of the model in these countries, it is not necessary to estimate resource costs, but only some parameters which have an equivalent function in the model. The purpose of the C parameters is to reflect the relative value of different resources; or conversely their relative scarcity or the relative difficulty of providing different resources. In a society with uniform and constant growths different resources are equivalently difficult to provide in the ratio of their current provision, and these ratios may be adequate first estimates of the C parameters. When different growths are expected of different parts of the HCS the ratios may be adjusted accordingly, or a more detailed analysis may reveal the "shadow prices" or each constrained resource.

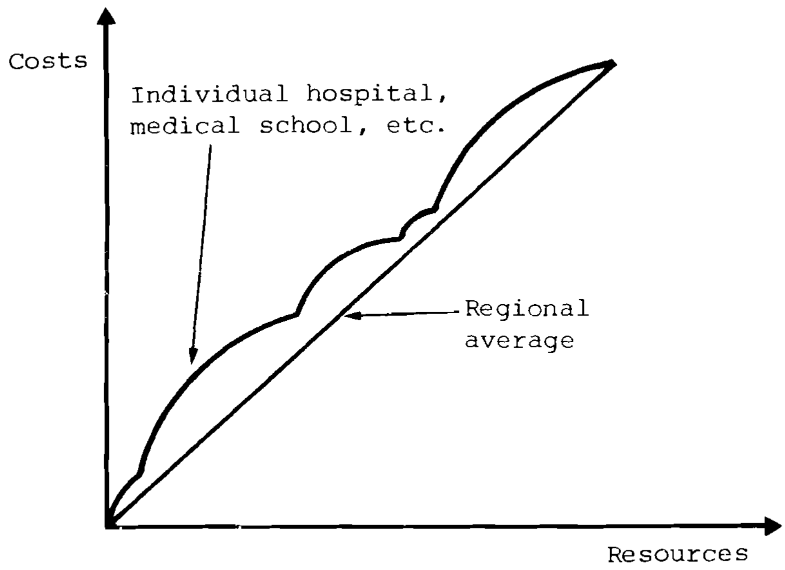

Figure 5. A linear regional cost function.

The principal outstanding problem is that of resource definition. The cost of a hospital bed could be the capital cost of creating it, or the revenue cost of maintaining it with food, heat and laundry. The cost of a doctor could include his training, his accommodation, or just his salary. The choices made at this stage actually *define* the resources for the purposes of the model, and they depend mainly upon which alternatives are interesting to the users of the model. Finally, of course, we really desire to estimate C at some future time instead of currently. A full treatment of this issue would need and could use more sophisticated predictive models.

4. ILLUSTRATIVE EXAMPLES

The procedures for model solution and parameter estimation described above have been implemented as computer programs. They are fairly compact, use no special software, and can be easily transferred to other computers. To illustrate their use and the use of the model, we present three illustrative examples of HCS resource allocation problems. The examples are extensions of those discussed previously in [3].

4.1 Example 1

The first example is designed to test the procedures for estimating α, β, X, Y. Consider the allocation of acute hospital bed-days in England between patients suffering from six diseases: varicose veins, haemorrhoids, ischaemic heart disease (excluding acute myocardial infarction), pneumonia, bronchitis, and appendicitis. Table 1 gives the numbers of patients admitted to hospital in 1973 with these diseases, and their average length of stay [9]. Gibbs used the empirical elasticities of Feldstein [8] and exogenous estimates of the ideal levels X and Y, to calibrate a predictive resource allocation model for these categories [2, 11].

Table 1. Example 1--actual allocations and model predictions.

	Allocation of hospital bed days[1)] in 1973 in England			
	Actual		Predicted by Model	
Disease	Admissions per 10 000 Population	Average Stay (Days)	Admissions per 10 000 Population	Average Stay (Days)
Varicose veins	7.6	10.1	7.6	10.4
Haemorrhoids	4.7	7.8	4.7	7.9
Ischaemic Heart	8.5	24.9	8.5	24.4
Pneumonia	14.0	18.0	14.1	18.0
Bronchitis	10.8	23.1	10.9	22.7
Appendicitis	17.5	7.9	17.5	7.9

1) 964.8 bed-days available per 10 000 population in 1973.

Here we repeat this exercise. However, we estimate the model parameters, not using Feldstein's figures, but by using the actual admissions and lengths of stay in the 16 health regions of England, in 1968 and 1973 [9,10]. Table 2 gives the parameters estimated by using the 1968 figures to estimate α, β, and the 1973 figures to estimate X,Y, recursively as described in Section 3.4. For this example, we have assumed that the parameters are constant over time, but this assumption could easily be relaxed.

The model parameters so estimated are different from those used by Gibbs [11] and also shown on Table 2, because of different data and because of the different values used for the arbitrary constraints. Nevertheless, they show very similar variations across diseases. Appendicitis is clearly represented as a disease where most patients must be hospitalised (high α), whilst bronchitis appears as a disease afflicting many patients (high X) but where hospitalisation is not essential (low α). The estimation procedure did not yield elasticities that were all positive, but those that were negative had so small an associated confidence that they could reasonably be changed to small positive numbers. The successful calibration of the model is confirmed by Table 1, which compares the model's prediction with the actual allocation of hospital bed-days in 1973 in England. The agreement is very close.

4.2 Example 2

The second example is the simplest possible example of a model with more than one resource. Table 3 shows the admission rates, lengths of stay, and doctoring received by patients in the 12 hospital regions of Czechoslovakia in 1975. The data is highly aggregated, including all hospital specialties, but it is potentially suitable for estimating model parameters for DRAM. Table 4 gives the results and shows that two of the three model elasticities can be adequately estimated but that the confidence associated with the third estimate is very small. Although these estimates might still be adequate for a predictive model, it would be better to try to improve them. Perhaps the data might be further disaggregated by category or by region. Alternatively, other years' data

Table 2. Example 1--two sets of model parameters.

Disease	Empirical Elasticities		Model Parameters				
	γ	η	α	β	X	Y	
Varicose Veins	0.78	0.62	1.64	3.03	12.8	15.4	In [11], α,β were
Haemorrhoids	0.70	0.44	2.11	4.68	7.7	13.1	estimated from
Ischaemic Heart	1.14	1.08	0.54	1.31	10.4	52.1	the data γ,η,
Pneumonia	0.71	0.23	2.28	9.87	21.0	19.7	and X,Y were
Bronchitis	1.13	-0.23	1.14	49.00	21.3	34.2	chosen exogen-
Appendicitis	-0.16	0.31	44.40	7.06	24.8	10.1	ously.
Varicose Veins	0.58	0.47	1.68	3.27	31.6	30.9	In Example 1, α,β and
Haemorrhoids	0.36	0.33	3.63	5.00	11.2	17.0	X,Y were estimated from
Ischaemic Heart	0.70	1.00	0.50	1.00	71.0	247.5	1968 and 1973
Pneumonia	0.71	0.19	1.57	9.44	75.2	28.1	regional allocations
Bronchitis	0.96	0.04	1.04	50.00	102.7	24.9	using $\phi = 5$, $\Theta = 20$.
Appendicitis	0.05	0.15	40.00	12.75	19.5	11.1	γ,η were then derived.

Table 3. Example 2--input data.

Region of CSSR	Figures for Czechoslovakia in 1975 — All hospital specialties: Population (thousands)	Admissions per thousand population	Average length of stay (days) (1)	Doctor days per admission (2)
Hl. m. Praha	1170	154.9	15.6	10.3
Středočeský	1136	188.1	13.4	4.6
Jihočeský	670	183.6	13.9	4.4
Západočeský	875	170.7	13.5	5.8
Severočeský	1140	179.7	13.7	4.3
Východočeský	1227	192.7	13.4	4.2
Jihomoravský	1992	169.6	13.9	5.1
Severomoravský	1883	185.7	14.2	4.7
Hl. m. SSR Bratislava	341	105.6	13.0	24.0
Západoslovenský	1636	147.1	13.9	4.5
Stredoslovenský	1462	149.7	12.9	5.4
Východoslovenský	1324	156.2	13.0	5.8
All CSSR	14857	168.4	13.8	5.5

(1) 2318.9 bed-days available per thousand population in 1975

(2) 931.3 doctor days available per thousand population in 1975

Table 4. Example 2--estimated parameters.

Costs: doctor:bed = 2.49:1.00 estimated from current availabilities assuming constant growth (see Section 3.5)

Ideal levels: $X = 400$

$Y_{beds} = 30$, $Y_{doctors} = 20$

arbitrary values, assumed to be known exogenously

Model Elasticities	Estimated Value (1)	Confidence (2)
α	0.69	0.9
β beds	1617.	-1.7
β doctors	12.5	0.9

(1) assuming $\phi = 50$ (2) Confidence $= \dfrac{\text{estimated value-standard error}}{\text{estimated value}}$

could be used to examine possible historical trends. We know now which of the parameters we need to investigate more thoroughly, and following such work, a health planner could use the model to investigate alternative policies for changing beds and staffing levels.

4.3 Example 3

The last example also considers the allocation of beds and doctors, but across two modes of care (inpatient and outpatient) and using data from the South Western Region of England. Table 5 presents historic allocation data from 1977 for the seven largest acute hospital specialties: general surgery, general medicine, obstetrics and gynaecology, trauma and othopaedic surgery, ENT, paediatrics, and ophthalmology [12,13]. In this example, the seven specialties are the patient categories, inpatient and outpatient are the two modes of care, and beds and doctors are the two resources. Therefore, this example uses all the structure available in the Mark 3 version of DRAM, although it has the simplifying feature that one of the resources (beds) is used in only one mode of care (inpatient).

Table 5. Example 3--historic resource allocations.

Hospital Speciality	Resource allocation in South Western RHA in 1977: Relevant catchment population	Admissions per thousand population: Inpatient	Admissions per thousand population: Outpatient	Average hospital stay (days) (7)	Half-day consultant sessions (5) (8): Inpatient (6)	Half-day consultant sessions (5) (8): Outpatient
General Surgery [1]	3035.4	20.9	19.0	7.87	0.170	0.153
General Medicine [2]	3035.4	14.8	10.5	10.18	0.183	0.345
Obstetrics and Gynaecology	1563.8 [3]	39.5	37.1	5.78	0.072	0.139
T & O Surgery	3035.4	9.1	22.4	13.60	0.252	0.121
ENT	3035.4	4.4	11.1	4.39	0.346	0.128
Paediatrics	641.8 [4]	29.7	17.7	6.28	0.266	0.362
Ophthalmology	3035.4	2.8	10.3	6.59	0.427	0.214

(1) Includes Urology
(2) Includes Cardiology
(3) Excludes males
(4) Excludes >15
(5) Assuming each full time consultant works the equivalent of 450 half-day sessions per year
(6) Derived by subtracting actual outpatient sessions
(7) 892.6 bed-days available per thousand population in 1977
(8) 50.87 half-day consultant sessions available per thousand population in 1977

Table 6 shows the model parameters which were estimated from similar data disaggregated for the five hospital areas of South Western, and available for 1976 and 1977. With only ten data points we would not expect to estimate a complete parameter set with great confidence, and some of the figures in Table 6 are very uncertain. Nevertheless, the variations between parameters are as expected. In obstetrics and gynaecology most of the demand is met (high α_j) but the need for outpatient treatment is very elastic (low β_{j22}). In general medicine the reverse is true. Many patients do not receive hospital treatment, but the supply of resources to those who do is rather inelastic.

Table 7 shows the predictions made by the model using these parameters for the resource levels actually used in 1977 and in Table 5. The agreement is not quite as good as for Example 1, because fewer data were available for calibration. Nevertheless, it demonstrates that sensible parameters can be estimated for the most sophisticated version of DRAM, even from relatively scanty data.

5. CONCLUSION

The user of DRAM Mark 3 is able to explore a wide range of planning issues. Not only may he study the consequences of changing the availability of different resources for patients in different categories, but also he may investigate how different modes of treatment compete for these resources in treating patients with different needs. The examples given above illustrate possible applications in acute inpatient treatment, but the model should be equally applicable in other care sectors, and perhaps even in other applications outside health care.

The next step in this work is to test and use the resource allocation submodel DRAM for applications in different countries. As our illustrative examples show, not all the structures modelled in DRAM need be used in every circumstance. Indeed, it is probably best to use the simplest possible formulation. Nevertheless, we have shown that a fully disaggregated DRAM can be both calibrated and implemented with only modest computing requirements.

Table 6. Example 3--estimated model parameters.

Hospital Speciality	α_j	X_{j1} IP	X_{j2} OP	β_{j11} IP B	β_{j12} IP D	β_{j22} OP D	γ_{j11} IP B	γ_{j12} IP D	γ_{j22} OP D
General Surgery[(1)]	10.0 (-0.4)	26.3	22.2	10.8 (0.7)	6.1 (0.7)	1.0 (-3.9)	10.5	0.34	0.46
General Medicine[(2)]	0.01 (0.5)	217.7	83.3	10.7 (0.3)	2.7 (0.7)	11.2 (0.8)	13.3	0.42	0.41
Obstetrics and Gynaecology	16.5 (0.2)	44.8	38.7	10.3 (0.6)	1.5 (0.8)	0.001 (0.6)	7.7	0.22	1.32
T & O Surgery	10.0 (-1.4)	10.8	26.5	1.0 (-1.4)	12.7 (0.2)	10.0 (-10.7)	58.5	0.37	0.15
ENT	10.0 (-0.7)	5.0	12.9	0.001 (0.2)	14.3 (0.0)	20.0 (0.7)	79.1	0.43	0.15
Paediatrics	5.6 (0.7)	43.7	19.4	8.9 (0.4)	5.8 (0.7)	1.0 (-4.8)	9.1	0.41	1.28
Ophthalmology	20.0 (0.0)	3.1	11.9	10.0 (-2.8)	8.3 (0.3)	10.0 (-2.3)	9.4	0.60	0.24

(1),(2) See Table 5
IP = Inpatient
OP = Outpatient

B = Beds
D = Doctors

Confidence coefficients as defined in Table 4 appear brackets

Table 7. Example 3--model predictions.

Hospital specialty	Predicted resource allocations [9] in South Western RHA in 1977					
	Relevant catchment population	Admissions per thousand population		Average hospital stay (days) [7]	Half-day consultant session per [8]	
		Inpatient	Outpatient		Inpatient	Outpatient
General Surgery [1]	3035.4	20.5	19.1	8.16	0.252	0.154
General Medicine [2]	3035.4	14.0	10.4	10.33	0.234	0.341
Obstetrics and Gynaecology	1563.8 [3]	38.3	36.2	5.91	0.090	0.148
T & O Surgery	3035.4	8.9	22.0	13.42	0.319	0.122
ENT	3035.4	4.4	10.7	4.18	0.371	0.137
Paediatrics	641.8 [4]	29.0	15.3	6.77	0.294	0.427
Opthalmology	3035.4	2.6	10.8	7.18	0.474	0.198

(1),(2),(3),(4),(7),(8) See Table 5.

(9) Using model parameters given in Table 6, and assuming relative costs of doctor:bed to be 1.57:1 as in [3].

APPENDIX 1

The Dual Optimisation Problem

Section 2 described a constrained optimisation problem. This appendix shows how the method used to solve this problem is equivalent to solving the dual optimisation problem by hill climbing.

We seek to maximise a function U(x,y) subject to the constraints

$$F_\ell(x,y) = \sum_j \sum_k x_{jk} y_{jk\ell} - R_\ell = 0 \qquad \forall \ell \qquad (4)$$

by using the Lagrangian

$$H(x,y,\lambda) = U(x,y) + \sum_\ell \lambda_\ell F_\ell(x,y) \quad . \qquad (5)$$

This primal problem is solved by finding

$$\max_{x,y} \min_{\lambda} H(x,y,\lambda) \quad .$$

The same values of x,y,λ, however, also solve the dual problem of finding

$$\min_{\lambda} \max_{x,y} H(x,y,\lambda)$$

in which equations (9),(10) are used to substitute for x,y, and an extremum is sought of H(x(λ),y(λ),λ).

We now observe that

$$\frac{\partial H}{\partial \lambda_\ell} = f_\ell$$

and

$$\left\{\frac{\partial^2 H}{\partial \lambda_\ell \partial \lambda_m}\right\} = \left\{\frac{\partial f_\ell}{\partial \lambda_m}\right\} = \underline{D}$$

in terms of the expressions in equations (14),(16),(17). Equation (15)

$$\lambda_\ell = \hat{\lambda}_\ell - \sum_m \bar{D}_{\ell m} f_m(\lambda) \qquad \forall\, \ell \tag{15}$$

is then revealed as a steepest descent algorithm for finding the minimum of H. The matrix

$$\underline{D} = \left\{\frac{\partial f_\ell}{\partial \lambda_m}\right\}$$

must be negative definite for convergence.

APPENDIX 2

Admissible Solutions of DRAM

Not all resource allocation patterns are admissible solutions of DRAM. This appendix describes the admissible solutions for the simplest possible DRAM with one patient category, one treatment mode and one resource.

As the resource level R (which is input to the model) increases, so also does the number of individuals treated, x, and the resource quota, y, allocated to each individual (which are outputs). For the simplest possible DRAM with J = K = L = 1, these two relations can be derived from equations (9) - (11) as

$$\left(\frac{R}{XY}\right) = \left(\frac{x}{X}\right)\left[\frac{\beta\left(\frac{x}{X}\right)^{-(\alpha+1)} + 1}{\beta + 1}\right]^{\frac{-1}{\beta}} \tag{A1}$$

$$\left(\frac{R}{XY}\right) = \left(\frac{y}{Y}\right)\left[\frac{(\beta + 1)\left(\frac{y}{Y}\right)^{-\beta} - 1}{\beta}\right]^{\frac{-1}{\alpha+1}} \tag{A2}$$

and it is easy to show that they have the shapes shown in Figures A1 and A2. Both curves are convex and monotonically increasing.

Alternatively, we may find an equation which relates x and y directly. The result

$$\beta\left(\frac{x}{X}\right)^{-(\alpha+1)} = (\beta + 1)\left(\frac{y}{Y}\right)^{-\beta} - 1 \tag{A3}$$

can have the three possible shapes shown in Figure A3. This figure gives the locus of solutions of DRAM on the xy plane. The particular solution for a given resource level is found at the intersection of the locus and the resource hyperbolae

$$\left(\frac{x}{X}\right)\left(\frac{y}{Y}\right) = \left(\frac{R}{XY}\right)$$

and it is the point on the hyperbolae which maximises the utility function of equation (1). Figure A4 depicts the shape of this utility surface above the xy plane.

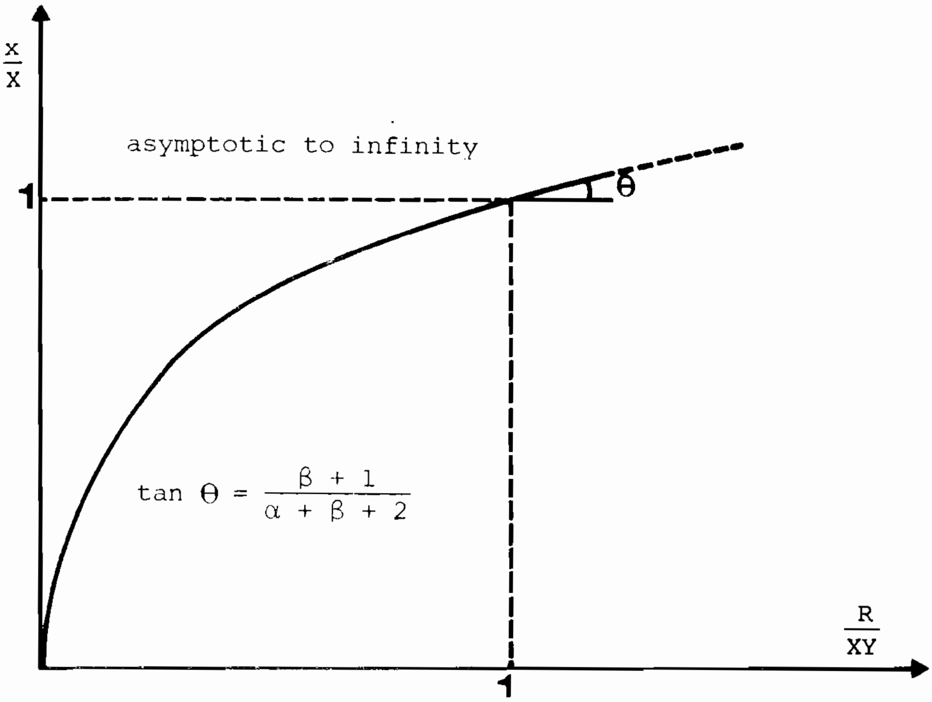

Figure A1. $\left(\frac{x}{X}\right)$ as a function of $\left(\frac{R}{XY}\right)$.

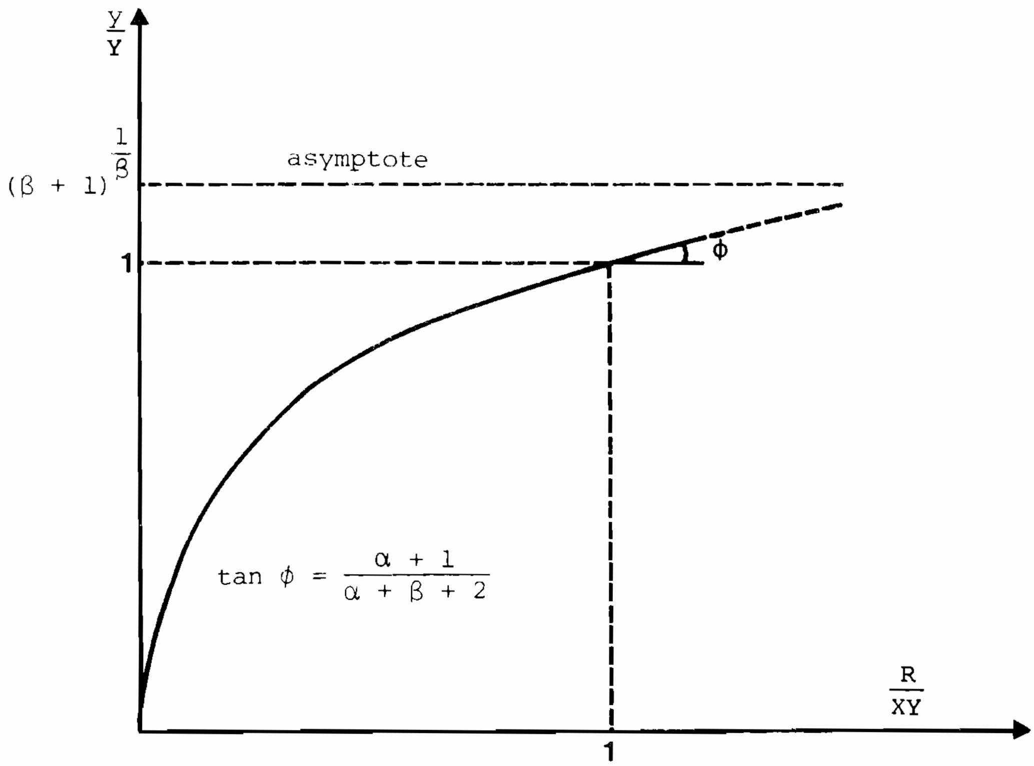

Figure A2. $\left(\frac{y}{Y}\right)$ as a function of $\left(\frac{R}{XY}\right)$.

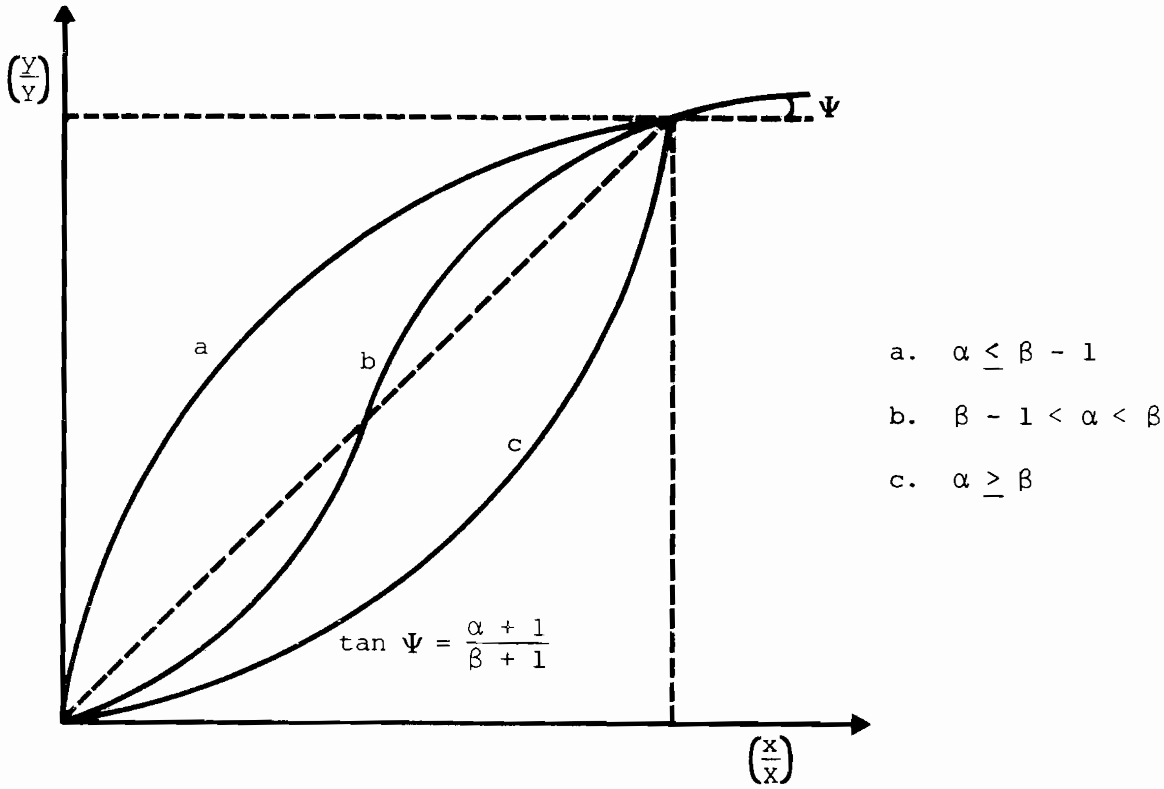

Figure A3. $\left(\frac{y}{Y}\right)$ as a function of $\left(\frac{x}{X}\right)$.

Finally, we consider what conditions two solution points

$$\left(\frac{x(1)}{X}, \frac{y(1)}{Y}\right) = (p_1, q_1) \tag{A4}$$

and

$$\left(\frac{x(2)}{X}, \frac{y(2)}{Y}\right) = (p_2, q_2) \tag{A5}$$

must satisfy if they are to lie on the same locus in Figure A3, and hence both to be admissible solutions of the same model. Substituting (A4),(A5) into equation (A3) and elimanating α gives

$$\left[\frac{(\beta + 1) q_2^{-\beta \log_{q_2} q_1} - 1}{\beta}\right] = \left[\frac{(\beta + 1) q_2^{-\beta} - 1}{\beta}\right]^{\log_{p_2} p_1} \tag{A6}$$

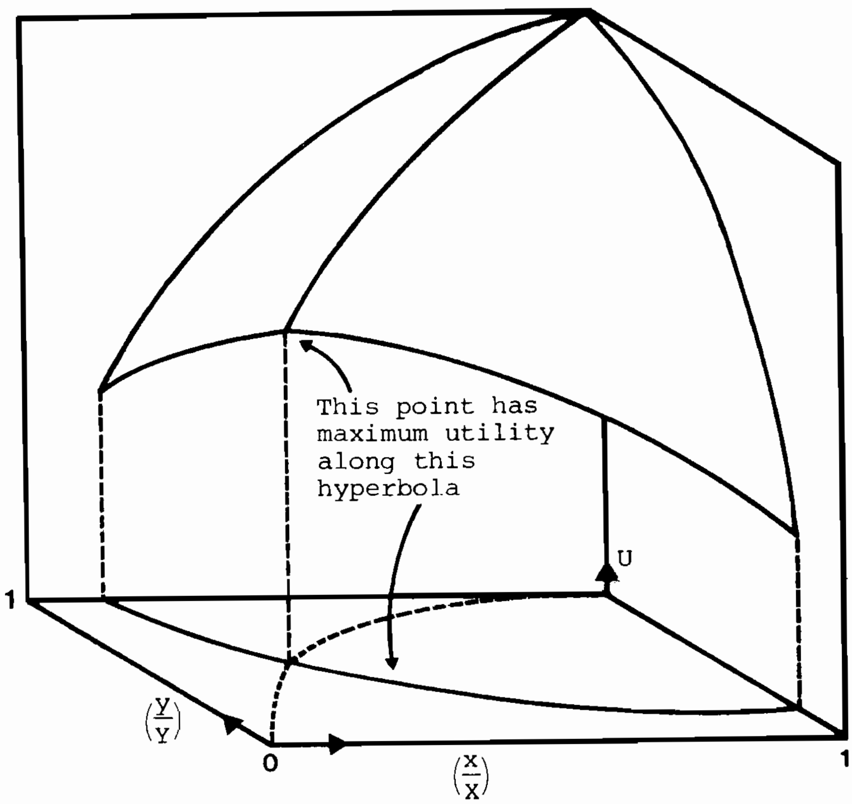

Figure A4. The shape of the utility surface above the xy plane.

which must be satisfied of all $\beta > 0$. Necessary conditions for this include:

a) $p_2 > p_1 \rightleftarrows q_2 > q_1$

b) if $p_2 > p_1$, $\log_{p_2} p_1 < \log_{q_2} q_1$.

Condition a) implies simply that an increase in the number of individuals treated, x, must be associated with an increase in the resource quota, y, allocated to them. It is not possible for one to increase and the other to decrease. Condition b) implies that the geometrical increase in $\left(\frac{y}{Y}\right)$ must always be greater than the geometrical increase in $\left(\frac{x}{X}\right)$. Equivalently, the diagonal in Figure A3 can be crossed only in an upward direction.

In versions of DRAM with more categories, modes, and resources, the conditions derived above will apply, together with other as yet unspecified conditions relating the different components of x and y. An interesting and unsolved question is the character of admissible solutions when the level of one resource increases and another decreases.

APPENDIX 3

Progress Towards a Combined Calibration and Validation Procedure

Section 3.1 outlined a procedure to estimate model parameters by minimising the difference in utility between the observed data and the corresponding model predictions. This procedure would be a comprehensive approach, but it would also be rather complicated and has therefore not yet been implemented. However, this appendix presents some further analysis.

Initially, we restrict ourselves to the simplest possible version of DRAM in which J = K = L = 1. The problem then is to choose α, β, X, Y so as to minimise ΔU_N of equation (21). Unfortunately, this problem is not convex, and additional constraints must be placed on the scale and shape of admissible solutions. A useful scale constraint is given by the ratio

$$\theta = \frac{XY}{x(1)y(1)} \tag{A7}$$

By reparameterising the problem in terms of Y and θ we can show that

a) $\left.\frac{\partial\lambda(i)}{\partial Y}\right|_{\theta \text{ constant}} = 0 \ \forall \ i$. In other words, constant θ implies constant $\lambda(i)$,

b) when θ is held constant, $\Delta U(i)$ and hence ΔU_N are strictly convex functions of Y. In other words, Y may be chosen so as to minimise ΔU_N.

There is no obvious shape constraint which has the same properties as these, but a possible candidate is the ratio

$$\phi = \frac{(\alpha + \beta + 1)}{(\alpha + 1)(\beta + 1)} \tag{A8}$$

This function is interesting because when $\beta \gg 0, \lambda \gg 1$, equation (19) reduces to

$$\lambda(i) = \left(\frac{XY}{x(i)y(i)}\right)^{\phi} \quad .$$

Under these conditions, therefore, we have a result similar to that obtained for Θ. Namely, that

$$\left.\frac{\partial\lambda(i)}{\partial\beta}\right|_{\phi \text{ constant}} = 0 \;\forall\; 1.$$

In other words, constant ϕ implies constant $\lambda(i)$. It seems likely that a corresponding convexity result also holds.

These results suggest the following approach to minimising ΔU_N.

a) Set values for Θ and ϕ, and using equations (A7),(A8) eliminate x and α from the expression (21) for ΔU_N.

b) Find by differentiation the gradients

$$\left.\frac{\partial\Delta U_N}{\partial Y}\right|_{\Theta,\phi \text{ constant}}, \quad \left.\frac{\partial\Delta U_N}{\partial\beta}\right|_{\Theta,\phi \text{ constant}},$$

c) Use improved hill-climbing methods to find the minimum of ΔU_N with respect to Y and β.

Whether this approach is useful depends upon whether it can be generalised to versions of DRAM which have many categories, modes, and resources. There is an obvious extended definition of Θ,

$$\Theta_\ell = \sum_j \sum_k X_{jk} Y_{jk\ell} \Big/ \sum_j \sum_k x_{jk}(1)\, y_{jk\ell}(1)$$

with the property that

$$\left.\frac{\partial\lambda_\ell(i)}{\partial Y_{jk\ell}}\right|_{\Theta_\ell \text{ constant}} = 0 \qquad \forall\; i \quad ,$$

but an equivalent ϕ_ℓ is not easy to formulate. Even if this problem can be overcome, improved hill-climbing methods will be inconvenient unless some decomposition between categories, modes, or resources is possible. ΔU_N is additive across these groupings, but the individual problems are still strongly coupled and possibly not individually stable.

APPENDIX 4

Unbiassed Regression Estimators

In the estimation of elasticities in Section 3.2, we assumed that elasticities are constant across the areas of a region, and then we performed regression analysis on the cross-sectional data. However, even if this assumption is incorrect and elasticities are different in different areas, we can show that this procedure still yields an estimate which reflects the regional "elasticity".

Define the indices:

j = area or subregion, $j = 1,2,\ldots J$

i = observation in each area, $i = 1,2,\ldots N$

and suppose that the data $x_j(i), y_j(i)$ satisfy the linear model

$$y_j(i) = b_j x_j(i) + \sigma\varepsilon_j(i) \tag{A9}$$

in which $\varepsilon_j(i)$ are uncorrelated random disturbances with zero mean. The unknown parameter b_j is different for different areas. Nevertheless, we assume that it is constant and form the usual least squares estimate

$$\hat{b} = \left(\sum_j X_j^T X_j\right)^{-1} \sum_j X_j^T Y_j \tag{A10}$$

in which $X_j = \{x_j(1),\ldots,x_j(N)\}^T$ and $Y_j = \{y_j(1),\ldots,y_j(N)\}^T$. We now investigate the properties of $\hat{b}$ when the unknown parameters b_j are assumed to be random samples from a Normal or Gaussian probability density function with mean m and variance v^2

$$b_j \sim N(m, v^2) \quad . \tag{A11}$$

Combining equations (A9) and (A10) gives

$$(\hat{b} - m) = \left(\sum_j X_j^T X_j\right)^{-1} \left\{ \begin{array}{l} \sum_j X_j^T X_j (b_j - m) \\ + \sum_j X_j^T E_j \sigma \end{array} \right\}$$

whence the results

$$E(\hat{b} - m) = 0$$

that the estimator $\hat{b}$ is an unbiassed estimator of the mean regional parameter m. Additionally, we may show that

$$E(\hat{b} - m)^2 \sim \frac{\sigma^2}{JN} + \frac{v^2}{J}$$

The first term on the right hand side is the usual residual variance term, while the second arises from the uncertainty about b_j.

REFERENCES

[1] Venedictov, D.D., et al, Health Care: A Systems Approach, in D.D. Venedictov (ed.), *Health Systems Modelling and the Information System for the Coordination of Research in Oncology*, CP-77-04, International Institute for Applied Systems Analysis.

[2] Gibbs, R.J., *The IIASA Health Care Resource Allocation Sub-Model: Mark 1*, RR-78-8, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1978.

[3] Hughes, David J., *The IIASA Health Care Resource Allocation Submodel: Mark 2--The Allocation of Many Different Resources*, RM-78-50, International Institute for Applied Systems Analysis, Laxenburg, Austria, September, 1978.

[4] Hughes David J., *The IIASA Health Care Resource Allocation Submodel: Formulation of DRAM Mark 3*, internal document, International Institute for Applied Systems Analysis, Laxenburg, Austria, October 1978.

[5] Kaihara, S., et al., *An Approach to Building a Universal Health Care Model: Morbidity Model of Degenerative Diseases*, RM-77-06, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1977.

[6] Fujimasa, I., S. Kaihara, and K. Atsumi, *A Morbidity Submodel of Infectious Diseases*, RM-78-10, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1978.

[7] Newhouse, Joseph P. and Charles E. Phelps, *Price and Income Elasticities for Medical Care Services*, R-1197-NC/OEO, The Rand Corporation, Santa Monica, California, USA, June 1974.

[8] Feldstein, M.S., *Economic Analysis for Health Service Efficiency*, North-Holland, Amsterdam, 1967.

[9] Department of Health and Social Security, UK, *Report on Hospital In-Patient Enquiry for the Year 1973*, HMSO, 1977.

[10] Department of Health and Social Security, UK, *Report on Hospital In-Patient Enquiry for the Year 1968*, HMSO, 1972.

[11] Gibbs, R.J., *A Disaggregated Health Care Resource Allocation Model*, RM-78-1, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1978.

[12] Department of Health and Social Security, UK, *Hospital Medical Staff Tables, 1976 and 1977*, unpublished.

[13] Department of Health and Social Security, UK, *SH3 Hospital Return Summaries, 1976 and 1977*, unpublished.

RELATED PAPERS OF THE HEALTH CARE SYSTEM MODELLING TASK

Venedictov, D.D., et al, Health Care: A Systems Approach, in D.D. Venedictov (ed.), *Health Systems Modelling and the Information System for the Coordination of Research in Oncology*, CP-77-04, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1977.

Kiselev, A., *A Systems Approach to Health Care*, RM-75-31, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1975.

Klementiev, A.A., *A Computer Method for Projecting a Population Sex-Age Structure*, RM-76-36, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1976.

Klementiev, A.A., *Mathematical Approach to Developign a Simulation Model of a Health Care System*, RM-76-65, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1976.

Shigan, E.N., *Alternative Analysis of Different Methods for Estimating Prevalence Rate*, RM-77-40, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1977.

Klementiev, A.A., *On the Estimation of Morbidity*, RM-77-43, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1977.

Fleissner, P., and A. Klementiev, *Health Care System Models: A Review*, RM-77-49, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1977.

Gibbs, R.J., *Health Care Resource Allocation Models--A Critical Review*, RM-77-53, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1977.

Gibbs, R.J., *A Disaggregated Health Care Resource Allocation Model*, RM-78-1, International Institute for Applied Systems Analysis, Laxenburg, 1978.

Gibbs, R.J., *The IIASA Health Care Resource Allocation Sub-Model: Mark 1*, RR-78-8, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1978.

Klementiev, A.A., and E.N. Shigan, *Aggregate Model for Estimating Health Care System Resource Requirements (AMER)*, RM-78-21, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1978.

Hughes, D.J., *The IIASA Health Care Resource Allocation Sub-Model: Mark 2--The Allocation of Many Different Resources*, RM-78-50, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1978.