UNCERTAINTY AND OPTIMAL POLICY: METHODOLOGY AND APPLICATION TO EDUCATIONAL FLOW MODELS

Ahmed Seifelnasr*

I. INTRODUCTION AND PURPOSE

In recent years the important role education plays in the developmental process has been recognized by many developing countries as indicated by the substantial increases in resource allocated to the education sector. Concurrently, national and international planning agencies have been giving serious attention to educational modelling with an attempt to incorporate the education sector into general socio-economic planning. In this paper we focus on some specific aspects in the use of educational flow models, in particular, how to deal with uncertainty regarding the values of some parameters and/or of some exogenous variables of a specified model.

Consider now the following typical stages in the educational planning process as described by Hooker (1974, p.83) and Werdelin (1972):

a) The present situation is analysed; past and present trends of development are studies and the future situation during a number of years is forecast;

b) targets for the future situation are assisted in various fields;

c) means of arriving at the targets are suggested; the feasibility of the steps is tested;

d) a system of follow-up is devised to check the development during the planning period and to allow new action whenever necessary. This process is continuous and iterative.

While the different planning approaches are mainly concerned with stage (b) above--that is, the target setting stage and differ in their conceptual basis in this regard--educational flow models are concerned with the internal dynamics of the educational system's supply. Thus, they are clearly indispensible and necessary tools in implementing the four stages mentioned above, especially stage (a). By estimating and tracing the flow of students and teachers throughout the educational system, educational flow model help to analyse and forecast educational supply. Stone (1965, 1966, 1968), Smith and Armitage (1967) and Alper, Armitage and Smith (1967) illustrate the development of computable student models. OECS (1970) and UNESCO (1966) give much more comprehensive flow models with applications.

Algebraically, a system flow may be summarized as follows in a matrix format:

$$\underline{X}(n) = A(n-1)\,\underline{X}(n-1) + \underline{B}(n-1) \qquad (1)$$

Where n : time

$\underline{X}(n)$: vector of dimension (rx1) where r is the number of states in the system (grades, for example). The i^{th} element, $x_i(n)$, is the number of students or teachers in state (i) at time n (the i^{th} state variable).

$A(n)$: the transition martix of the system of order (rxr). The elements of this matrix are the transition coefficients that describe the movement within and out of the system. Of course not all transitions between the various states are feasible. This implies that many of the

: elements of the transition matrix are zeros

$\underline{B}(n)$: A (rxl) vector of new entrants into the system between (n) and (n+1).

The matrix equation (1) represents the basic model of flow and supply dynamics and it is now the generally accepted format in most supply forecasting (Davis, 1974).

For the purpose of this paper, educational modelling can be divided into two stages. The first has to do with building a model that is appropriate for the policy questions being examined and the second has to do with developing appropriate current and future values of the parameters and/or exogenous variables. It is this second stage that constitutes the frame of reference for the discussion and methoology of this paper.

It is easy to see that setting aside the possibility that the structure of the flow model is a bad one, the validity of the outcome will depend largely on the various given to the parameters and exogenous variables. But, since the parameters setting is carried out exogenously, the flow model by itself cannot produce realistic forecasts. Special attention, therefore, should be given to possible variations in the values of exogenous variables and parameters estimated in the parameters setting stage. These estimates are not necessarily certain. They are usually accompained by a range of estimation errors. This uncertainty is compunded by the accumulation of such errors when the values of the exogenous variables are projected into the future. On statistical grounds the estimated values of the exogenous variable are subject to some given deviations.

Whatever the approach to educational planning (social demand, manpower requirements, etc.), the objective of the plan is usually transformed to quantitative targets to be reached over a given period of time (the planning period). We will consider the target which is expressed in terms of desired number and distribution of students and/or teachers in different educational levels.

To choose between the different possible alternatives that can be generated using the educational flow models of how these targets are to be achieved, it is assumed that the planner will choose that one which optimizes a given objective function. This process is bound to be affected by the possible variations of the values of parameters and/or exogenous variables discussed above. The purpose of this paper is to develop some analytical and operational tools that can be used to provide an answer to the following question:

How does a given change in the value of some or all parameters and/or exogenous variables in educational flow models (as represented by Eq. (1)) affect the final results of the optimization process?

It must be emphasized that the purpose of these tools is not to eliminate the elements of uncertainty regarding the true values of those parameters and/or exogenous variables. It rather attemtps to analyse systematically systematically the likely consequences of their presence and in this sense reduces the impact of uncertainty in planning.

II. THE PROBLEM

Within the context of the above discussion and to state the problem in a more general form, let the educational flow model be represented by the following set of equations:

$$\underline{X} = \underline{f}(X(o), \underline{P}, \underline{d}, \underline{u}) \qquad (2)$$

where $\underline{X}$ is the vector of state variables, $\underline{P}$ is a vector containing the model's parameters, $\underline{d}$ is a vector containing the exogenous (or input) variables, $\underline{X}(o)$ is the initial condition vector, and $\underline{u}$ is a vector of policy or decision variables. Those last variables are assumed not to be generated by the model, but they can be parameters and/or exogenous varaibles which are subject to influence or manipulation by the planners.

We also add the set of equations:

$$\underline{Y} = \underline{g}(x) \qquad (3)$$

where $\underline{Y}$ is the vector of output variables. Of course, the function $\underline{g}$ could be the identify function if the output is the levels of the state variables themselves. However, if we are interested in some and not all state variables, their proportional distribution, etc., then Eq. (3) has to be augmented to the model.

Let the planning period be N years. To carry out a simulation run over this period, the planner has to supply a sequence of values for the vectors $\underline{P}$, $\underline{d}$, and $\underline{u}$ for the entire planning periods in addition to the initial conditions. The simulation then produces an output trajectory ($\underline{Y}$).

Let (Y) be the desired trajectory of output variables (which are determined in the target setting state mentioned above). Because of political and economic constraints or because of the importance of some policy variables by themseleves (i.e., not only because of their effect on output variables), e.g., in the case of the desire to decrease the wastage in education by reducting the dropout rate or when there is an attempt to improve the quality of education by increasing the teacher-student radio, etc., we will assume that there is a similar desired path for policy variables ($\hat{\underline{u}}$) to which the planner hopes to stay close.

Assuming that the planner has chosen the following quadratic function to represent the preference function mentioned above:

$$J(Z) = Z' \, WZ \qquad (4)$$

where $$Z = \begin{bmatrix} \underline{Y} - \underline{Y} \\ \underline{u} - \underline{u} \end{bmatrix}, \text{ of order } (n_2 \times n_1) \times 1;$$

$n_1 = (m_1 \times N)$, $n_2 = (m_2 \times N)$, where m_2, m_1 are the number of output and policy variables defined for each period of time. The weighting matrix is a diagonal matrix or order $(n_1 + n_2) \times (n_1 + n_2)$ with weights assumed to reflect the priorities of the planner for the importance of a group of variables versus the others. The planner is interested in choosing the time path of the policy variable which minimizes this quadratic function.

The planner's concern now, given the problem just defined, is that he (or she) is not certain about the value taken by certain numbers in the model (parameters and/or predicted exogenous variables). Regardless of the role these numbers play in the model, they will be called error terms and will be put together in a vector $\underline{\epsilon}$ with n_3 components. Let $\bar{\epsilon}_i$ be the nominal value assigned to the i^{th} component of the vector $\underline{\epsilon}$ during simulation and optimization. Based on experience or by examining past data, we assume that the uncertain knowledge of the component ϵ_i can be expressed by stating that could take any value within the range $\epsilon_i \pm e_i$.. This range will be referred to as the range of uncertainty. Given a similar statement for each component of the vector $\underline{\epsilon}$, such statements will generate a set that surrounds the nominal vector $\bar{\epsilon}$ and will contain all the possible joint realisations of the different components. This set will be referred to as the error set E.

The question which was posed in the previous section can now be broken down into two questions:

1) If the optimum values were calculated using the nominal values $\bar{\underline{\epsilon}}$, how much would other possible values of $\underline{\epsilon}$ change the optimum?

2) What is the image of the set E in policy space and what are its characteristics?

In this paper an attempt is made to answer these two questions.

III. METHODOLOGY

Two groups of measures from the basis for the analysis. The first group depends on the particular relationships and definitions specified by the two equations (2) and (3) together with the numerical specifications of the parameters and initial condition. These specifications will be referred to collectively as system structure. The second group of measures depends on some measures of the first group together with the range of uncertainty assumed for the components of the vector $\underline{\epsilon}$.

A. Measures that Depend on System Structure

If all the information contianed (or required) in Eq. (2) is known and fixed except regarding the required values of policy variables and the uncertain values in $\underline{\epsilon}$, then the output variables $\underline{Y}$ can be considered as a function of both $\underline{u}$ and $\underline{\epsilon}$:

$$\underline{Y} = \underline{Y}(\underline{u}, \underline{\epsilon}) \tag{5}$$

The set of equations given by (5) can be nonlinear equations, and, depending on the definition of output variables and the relationship assumed in the model, it may even be difficult to write the R.H.S. of this equation in explicit form. However, for a small enough change, $\delta \underline{u}$ to $\underline{u}$ and $\delta \underline{\epsilon}$ to $\underline{\epsilon}$, a linear relationship is assumed to hold between change $\delta \underline{Y}$ and $\delta \underline{u}$ and $\delta \underline{\epsilon}$. Thus:

$$\delta \underline{Y} = S_1 \, \delta \underline{u} + S_2 \, \delta \underline{\epsilon} \qquad (6)$$

with the matrices S_1 and S_2 has obvious interpretation.

Define $S_u = \begin{bmatrix} S_1 \\ \hline I \end{bmatrix}$ and $S_\epsilon = \begin{bmatrix} S_2 \\ \hline 0 \end{bmatrix}$

where I is a unit matrix of order $(n_1 \times n_1)$ and 0 is a zero matrix of order $(n_2 \times n_3)$. Thus we have :

$$\delta \underline{Z} = S_u \, \delta \underline{u} + S_\epsilon \, \delta \underline{\epsilon} \qquad (7)$$

The optimization problem here is one of the so-called unconstrained optimization problems and can be solved using one of the algorithms designed for this class of problems, particularly when the objective function is a sum of squares of functions, e.g. Gauss-Newton or its modified versions (see Murray, 1972).

All these algorithms are iterative and they calculate a sequence of points $\underline{u}^{(1)}$, $\underline{u}^{(2)}$, ..., starting from initial point $\underline{u}^{(0)}$, that should converge to a point $\underline{u}^*$ that minimizes the objective function. These algorithms calculate $\underline{u}^{(i+1)}$ from linear approximation to the function $\underline{Z}$ (using our notation), this approximation being chosen so that they are good near $\underline{u}^{(i)}$. In our case the amount of increment in the function $\underline{Z}$ will be provided by equation (7).

In these algorithms the amount of correction ($\delta\underline{u}$) to the vector $\underline{u}$ during the iteration is the one which minimizes the objective function $J(\underline{Z} + \delta\underline{Z})$ which, using equation (7), is now given by:

$$\delta\underline{u} = -(S_u'WS_u)^{-1} \; S_u'W(\underline{Z} + S_\epsilon \delta\underline{\epsilon}) \qquad (8)$$

Defining $\delta\underline{\epsilon} = \underline{\epsilon} - \bar{\underline{\epsilon}}^{1}$, we could compue the nominal optimal policy vector, which will be demoted by $\bar{\underline{u}}^{*}$, by letting $\delta\underline{\epsilon} = 0$ (i.e., assuming there is no uncertainty regarding the value of the components of the vector $\underline{\epsilon}$) with the new iteration beginning with the new value of the vector $\underline{u}$ computed by:

$$\underline{u}_{new} = \underline{u}_{old} + \delta\underline{u}$$

Convergence is assumed when δu is small enough. From the nominal optimum $\bar{\underline{u}}^{*}$, the change in u^* due to a given change $\delta\underline{\epsilon}$ is from equation (8):

$$u^*(\delta\epsilon) = \bar{u}^* + \Delta\underline{u}(\delta\epsilon), \qquad (9)$$

$$\Delta\underline{u}(\delta\epsilon) = -(S_u'WS_u)^{-1} \; S_u'WS_\epsilon = -C \; S \; S = -Bo \qquad (10)$$

The effect of specific deviations in the error terms on the variables of interest in the model can be examined through two measures. The first measure comes from the matrix S_ϵ (or S_2) which is the matrix of the sensitivity coefficients $\left(\frac{\partial Y}{\partial \epsilon}\right)$. The columns of this matrix give the individual effect of a unit deviation in the error terms on the target variables around their optimum values and it is easily constructed by simulating with the nominal optimum policy ($\bar{\underline{u}}^{*}$) and small change in turn in what constitutes the error terms $\epsilon_1, \epsilon_2, \ldots\ldots, \epsilon_{n_3}$. The second measure comes from the matrix B defined in equation (10). The (i^{th}) column of this matrix gives the shift in the optimum policy resulting from a unit deviation in the (i^{th}) error term.

B. Measures that Depend on System Strucutre and the Range of Uncertainty

We notice that altough matrix B gives the effect of deviation in error terms on policy variables, it only gives the effect of each one separately. It does not provide joint effects. It is undoubtedly more interesting and more important to planners to examine the joint effect of a specific combination of deviations versus the joint effect of a different combination. In addition, when comparing the effect of different "groups" of errors a more accurate and complete comparison is the one which compares both ~~the joint and~~ the individual effects A first step here is to define the set E.

From among the possible definintions of the set E we will conside only one, namely the ellipsoidal set E:

$$E \ : [\underline{\epsilon} : \delta\epsilon' V^{-1} \delta\underline{\epsilon} \leq 1] \qquad (11)$$

Where $\delta\underline{\epsilon} = \underline{\epsilon} - \bar{\underline{\epsilon}}$

and V is a diagonal matrix of order ($n_3 \times n_3$) with e_i^2) as its i^{th} diagonal element. Inclusion of off diagonal terms in V gives a rotated ellipsoid and interacting errors.

Assuming that all possible combinations of the component of the vector $\underline{\epsilon}$ are equally likely to occur is equivalent to assuming that the vector $\underline{\epsilon}$ is uniformly distributed over the set E .[2] This leads us to the concept of ellipsoid of concentration of ϵ^{3}. Accordingly, the density defined by a uniform distribution over the interior of the ellipsoid:

$$\delta\underline{\epsilon}' V^{-1} \delta\underline{\epsilon} = n_3 + 2 \qquad (12)$$

has a mean vector equal to $\bar{\epsilon}$ and a covariance matrix equal to V.

The linear relationship defined by equation (10) defines an updating vector $\Delta\underline{u}(\delta\underline{\epsilon})$ for each $\underline{\epsilon}$ (and hence for each $\delta\underline{\epsilon} = \underline{\epsilon} - \bar{\underline{\epsilon}}$) in the set E now defined by:

$$E : [\underline{\epsilon} : (\underline{\epsilon} - \bar{\underline{\epsilon}})' \frac{1}{n_3+2} V^{-1} (\underline{\epsilon} - \bar{\underline{\epsilon}}) \leq 1] \qquad (13)$$

From equation (10) another set containing all the possible updated vectors $\underline{u}^*$ can be defined as follows:

$$U : (\underline{u}^* : \underline{u}^* = \bar{\underline{u}}^* + \Delta\underline{u}(\delta\underline{\epsilon}) ; \underline{\epsilon} \text{ in } E) \qquad (14)$$

The policy vector $\underline{u}^*$ is now a randmm vector since the updating vector $\Delta\underline{u}(\delta\underline{\epsilon})$ is now a function of the random deviations $\delta\underline{\epsilon}$ through equation (10). Thus $\underline{u}^*$ is a n_1 -components random vector with:

$$E(\underline{u}^*) = \bar{\underline{u}}^* \qquad (15)$$

$$Cov(\underline{u}^*) = E[(\underline{u}^* - \bar{\underline{u}}^*)(\underline{u}^* - \bar{\underline{u}}^*)'] = E(\Delta\underline{u}\Delta\underline{u}') = \frac{1}{n_3+2} BVB' \qquad (16)$$

It then follows that the ellipsoid of concentration of $\underline{u}^*$ is given by:

$$\Delta\underline{u}' (Cov(\Delta\underline{u}))^{-1} \Delta\underline{u} = n_1 + 2$$

$$\Delta \underline{u}' \left[\frac{BVB'}{n_3 + 2} \right]^{-1} \Delta u = n_1 + 2$$

$$\Delta \underline{u}' \left[\frac{n_1 + 2}{n_3 + 2} BVB' \right]^{-1} \Delta \underline{u} = 1 \qquad (17)$$

This ellipsoid encloses all possible combinations of the components of the vector $\underline{u}^*$ which correspond to the possible combinations of components of the deviation vector $\delta \underline{\epsilon}$; both combination are uniformly distributed over the interior of the corresponding ellipsoid.

Some basic results from linear algebra[4] can be used to characterize the set U given by:

$$\text{U : (}\underline{u}^* : \Delta \underline{u}' V_1^{-1} \Delta \underline{u} \leq 1; \epsilon \text{ in E)} \qquad (18)$$

where $V_1 = \frac{n_1 + 2}{n_3 + 2} \cdot BVB'$.

The dimension of the set U is given by the rank of the matrix BVB' which has the same rank as the matrix B defined in equation (10) Thus:

$$\text{Dim (U)} = \text{rank (B)} = \min (\text{rank } (S_u), \text{rank (W)}, \text{rank } (S_\epsilon)) \leq \min (n_1, n_3).$$

U may be confined to a subspace of lower dimension than n_1 if there are fewer than n_1 errors, for example.

The shape of U can be easily determined from the directions and lengths of the axes of the ellipsoid

$$\Delta \underline{u}' V_1^{-1} \Delta \underline{u} \leq 1$$

which are found from the eigenvalues and eigenvectors of the symmetric positives semidefinite matrix V_1. The j^{th} axis has length $2\ell_j$, where ℓ_j is the square root of the j^{th} eigenvalue and direction $v^{(j)}$, where $v^{(j)}$ is the eigenvector corresponding to the j^{th} eigenvalue and of unit length.

If U is of lower dimension than n_1, some eigenvalues are zeros and the nonzero axes are an othonormal basis for U. This means that every vector in the set U (i.e., every updated policy vector) can be expressed as a linear combination of them. Thus the eigenvalues and eigenvectors give the length and direction of change within the set U.

The measures intorduced to characterize set U help to analyse the joint effect of different errors. These measures are the eigenvalues and eigenvectors of the matrix V_1 defined in equation (18) showing the directions and lengths of major patterns of variation within the set U, and the variance covariance matrix of the vector $\Delta \underline{u}$, Cov($\Delta \underline{u}$), given by equation (16).

IV. FINDINGS : ILLUSTRATIVE EXPERIMENTS

To illustrate the concepts discussed in the previous sections, five experiments (a-e) were performed. In all experiments the main interest was to see how the different measures aimed at examining and quantifying the dynamic effects of errors reflect the basic differences between the experiments. In all experiments the same definintion of policy variables was used. The experiments (a), (b) and (c) are the same respect except that they differ in what constitutes the source of uncertainty, i.e., the components of the vector . The last two experiments (d) and (e) have the same source of uncertainty as experiment (b), but they differ from the rest of the experiments and between themselves in their definition of target (output) variables.

The model used in all the experiments to specify the relationships between the state variables is a simple 2-sector model of teacher supply:

$$X_1(n+1) = X_1(n)a_1 + g_1(n)\,a_4$$

$$X_2(n+1) = X_2(n)a_3 + X_1(n)a_2 + g_2(n)a_5$$

where

$X_1(n)$: number of teachers of type (1) at time (n)

$X_2(n)$: number of teachers of type (2) at time (n)

$g_1(n)$: number of graduates of teacher training college for type (1) at time (n).

$g_2(n)$: number of graduates of teacher training college for type (2) at time (n)

a_4 : proportion of type (1) graduates who enter the teaching profession.

a_1, a_3 : retention ratios for type (1) and type (2) respectively

a_2 : proportion of type (1) teachers who become type (2) teachers.

Those relationships provide the supply dynamic of the national teachers only. We assume that there are a number of foreign teachers in the system at the beginning of the planning interval whose future survivors in the system were estimated exogenously. If we assume further that the planner is interested in the total number of teachers available in the system (nationals plus foreigner) at each year, then the output variable is defined to be this tital. The desired yearly values for this output variable are, of course, the total annual requirements for teachers in the system. Those total requirements were assumed to be computed exogenously and given to the planner as fixed in all experiments (but not necessarily constant). In computations, the variables that appear in the objective function were defined as to have a desired value of zero.[5] Thus the "modified" target varibale was defined as follows for each time period:

$$Y(n) = [\, T(n) - D(n) \,] \qquad (19)$$

where $T(n)$: total number of teachers at time (n)

$D(n)$: total requirements of teachers at time (n).

Two policy variables were defined for each time period. Taking the last observation into account, these were:

$$u_1(n) = \frac{a_5(n)}{a_5(n-1)} - 1.035 \qquad (20)$$

and $$u_2(n) = \frac{I(n)}{g(n)} - .3 \qquad (21)$$

where $a_5(n)$: the value of the proportion of type (2) graduates who enter the teaching profession at time (n) with $a_5(0) = a_5$

I(n) : number of foreign teachers imported at time (n)

g(n) : total number of graduates of both types who enter the teaching profession at time (n).

Thus, equation (20) means that the first variable deemed controllable by the planner is the proportion (a_5) with a desirable annual rate of increase of .035. In equation (21) it was assumed that one objective of the planner is to control the number of foreign teachers coming to the system at time (n) such that it may not exceed one-third (or a little less) of the total number of the nationals who enter the system at time (n).

The total number of teachers defined in quation (19), therefore, has to include those foreign teachers and their survivors[6] in the ~~system as~~ well. Thus, the total number of teachers is defined as follows:

$$T_{(n)} = X_1(n) + X_2(n) + I_0(n) + I(n) + I_S(n) \qquad (22)$$

where $I_0(n)$: the number of foreign teachers who were in the system at the beginning of the planning interval and who are still in it at time (n)

I(n) : the number of foreign teachers who come to the system during time (n)

$I_S(n)$: the number of foreign teachers who came to the system before time (n) and who are still available at time (n).

The length of the planning interval (N) was assumed to be 10 years. But, since equations (20), (21), and (22) specify that values for the target variables and for each of the two policy variables are to be determined each year, the total number of target variables and policy variables were 10 and 20 respectively in each experiment .

The aim of the optimization is to minimize a weighted sum of squares of all 30 variables:

$$J = \sum_{i=1}^{30} w_i z_i^2 = \sum_{i=1}^{10} w_i y_i^2 + \sum_{i=1}^{20} w_{i+10} u_i^2 . \qquad (23)$$

The weights were chosen somewhat arbitrarily as follows:

$w_i = 1,\ i = 1, 10$

$w_j = 1 \times 10^5,\ j = 11, 13, 15, \ldots, 19$ (those associated with the first policy varibale)

$w_j = 1 \times 10^2,\ j = 12, 14, \ldots, 20$ (those associated with the second policy variable).

It is to be emphasized that our main concern here is not the optimum solution that satisfies equation (23) (referred to hereafter as the nominal optimum) per se, but in "using" it as a base-point for the dynamic sensitivity analysis.

In the fourth experiment (d) the planner was assumed to be concerned with the path of the first differences (instead of the levels) of the target variables defined by equation (20). Thus, in this experiment equation (20) was replaced by the following equation:

$$Y(n) = ([T(n) - T(n-1)] - [D(n) - D(n-1)]) \qquad (24)$$

Apart from this change, experiment (d) is similar to experiment (b) including the definition of the error terms.

In the last experiment (e) the definition of target variable changed again, this time to a nonlinear function of the state variables. It was assumed that the planner is interested only in the proportional distribution of the state variables and he would like to see this distribution follow a given path. The target variable now has the form:

$$Y(n) = \frac{\hat{X}_1(n)}{\hat{X}_1(n) + X_2(n)} - C(n) \quad (25)$$

where $\hat{X}_1(n) = X_1(n) + I(n)$ (26)

and C(n) : desired path of the ration of number of teachers type (1) to total teachers with C(0) = .82 and C(10) = .60

Thus it was assumed that the foreign teachers who would come to the system at time (n) would be of type (1). This new definition nec necessitates a change in the weights associated with the target variables which now become:

$$w_i = 1 \times 10^3,\ i = 1,2....10.$$

The initial conditions and the numerical values of the parameters are as follows:

$X_1(0) = 1400$, $X_2(0) = 300$, $a_1 = .85$, $a_2 = .05$, $a_3 = .75$, $a_4 = .85$, $a_5 = .65$, $g_1(0) = 350$, and $g_2(0) = 60$

The error terms and their assumed ranges in the different experiments are as follows:

Experiment	Error Term	Nominal Value	Range ($\pm e$)
(a)	$x_2(0)$, $x_1(0)$	300, 1400	$\pm$ 16, $\pm$ 124
(b)	a_1 , a_3	.85, .75	$\pm$.05, $\pm$.075
(c)	a_4 , a_3	.85, .75	$\pm$.05, $\pm$.075
(d)	as (b)	as (b)	as (b)
(e)	as (b)	as (b)	as (b)

A final and important observation is in order. From the assumption of the model and from the definition of both policy variables and target variables, it was implicitly assumed that a decision taken at time (n) will affect only the state and target variables at time (k) when $k \geq n$. This assumption, which is a natural one, has important implications.

Recall equation (6) :

$$\delta \underline{y} = S_1 \delta \underline{u} + S_2 \delta \underline{\epsilon}$$

which gives the first order effect of change in policy variable ($\delta \underline{u}$) and in error terms ($\delta \underline{\epsilon}$) on target variable ($\delta \underline{y}$).

The matrix S_1 is of order (10x20) since we have one target variable and two policy variables for each year. The assumption mentioned above implies that S_1 can be partitioned into 10^2 sub-matrices all of order (1x2) as follows:

$$S_1 = \begin{bmatrix} S_{1,1} & 0 & 0 & 0 \\ S_{2,1} & S_{2,2} & 0 & 0 \\ S_{3,1} & S_{3,2} & S_{3,3} & 0 \\ \\ S_{10,1} & S_{10,2} & S_{10,3} & S_{10,10} \end{bmatrix} \qquad (27)$$

The diagonal blocks (S_{11}, S_{22},...) specify the partial derivatives of the target variable with respect to policy variables in the same per. period, and the submatrices below and to the left of these (S_{21}, S_{31}...) specify the partial derivatives of target variable with respect to policy variables in earlier periods. This particular form of this matrix has important implications which will be discussed later.

V. CONCLUSION

In this paper an attempt was made to examine and analyse the effect of errors on the values taken by some variables in educational planning models . Five experiments were attempted; the results lead to the following observations:

a) In this type of analysis the numerical specifications of the parameters and the initial conditions in the model play the dominant role in determining the relative importance (in terms of greater effect) of one error term versus another. Figure 1 shows the individual effect of a 5% change in the error terms on the target variable in experiments (b) and (c), namely, a_1, a_3, and a_4. From

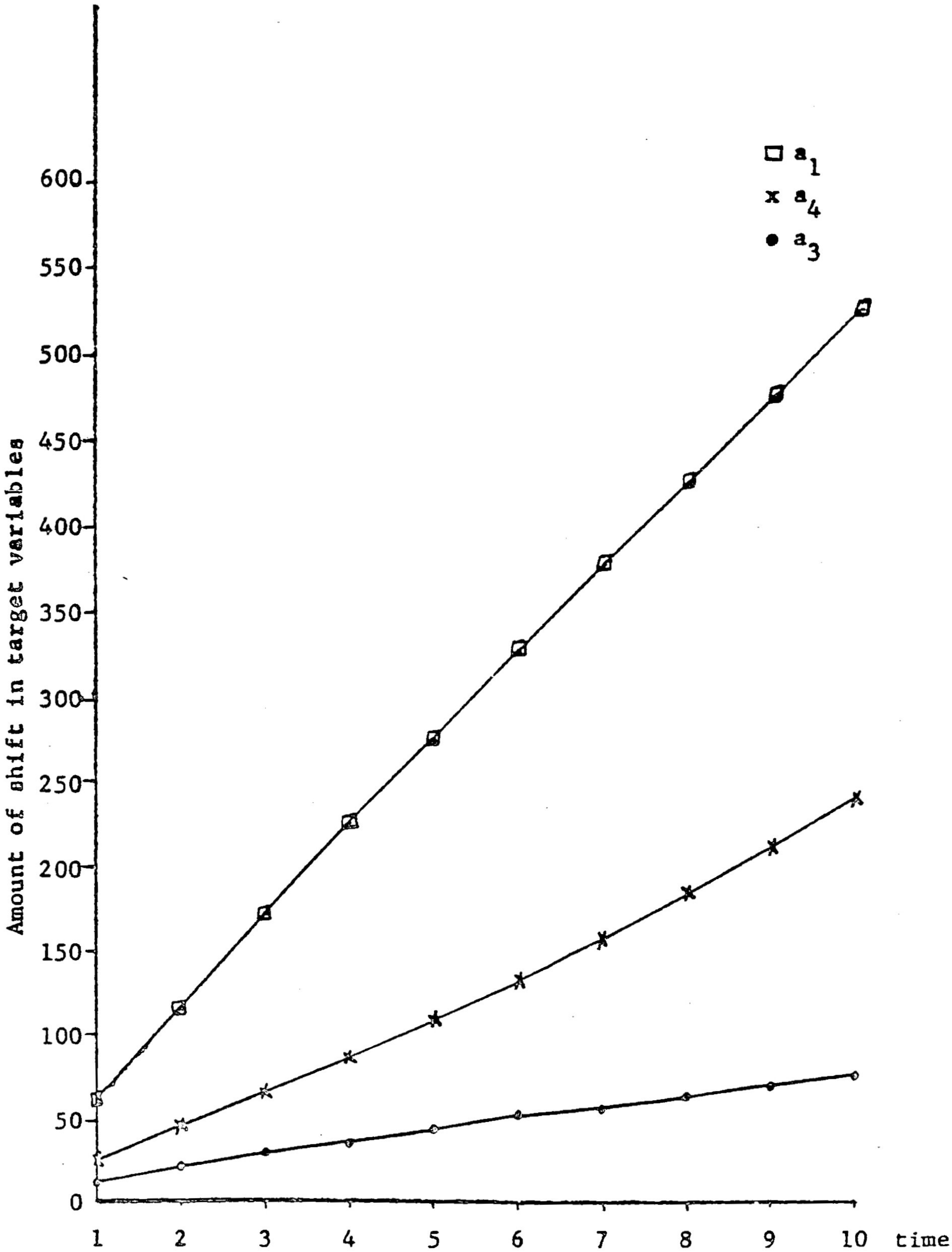

Figure (1). The effect of errors in parameters a_1, a_3 and a_4 (equivalent to 5% of their nominal values) on target variables.

the figure it is clear that a change in a_1 (associated with the larger of the two state variables) has the largest effect on the target variables followed by a_4 and a_3.

b) The effect of a deviation in an error term on output variables depends totally on how these variables are defined as functions of the state variables. The reason is that the observed change in output variables is a "net" result of two effects. The first effect is the effect of change in the error terms on the state variables and the second is the effect of the change in the state variables on the output variables. If those two changes move in the same direction[8] (i.e., an increase [decrease] in state variables which results finally in an increase decrease in output variables), the net result on output variables will be amplified. If the two changes move in opposite directions[9], they may balance out each other and the net effect on output variables may be smaller than expected.

c) All the measures intorduced are functions of two matrices (which are measures by themselves): the matrix S_ε which measures the sensitivity of output variables to individual deviations in the error terms, and matrix B which measures the effect of these deviations on the policy variables. To compare the effect on target variables of different magnitude, the elements of the matrix S_ε were normalized by dividing them by the optimal value of the corresponding target variable. Figure 2 shows the normalized sensitivity coefficient in experiments (b) and (e). This figure confirms the significance of observations (a) and (b) above. The relative importance of a_1 versus a_3 in experiment (b) completely reversed in experiment (e). Thus a unit deviation in a_3 will have more effect (relative to the optimum) than a corresponding devia-

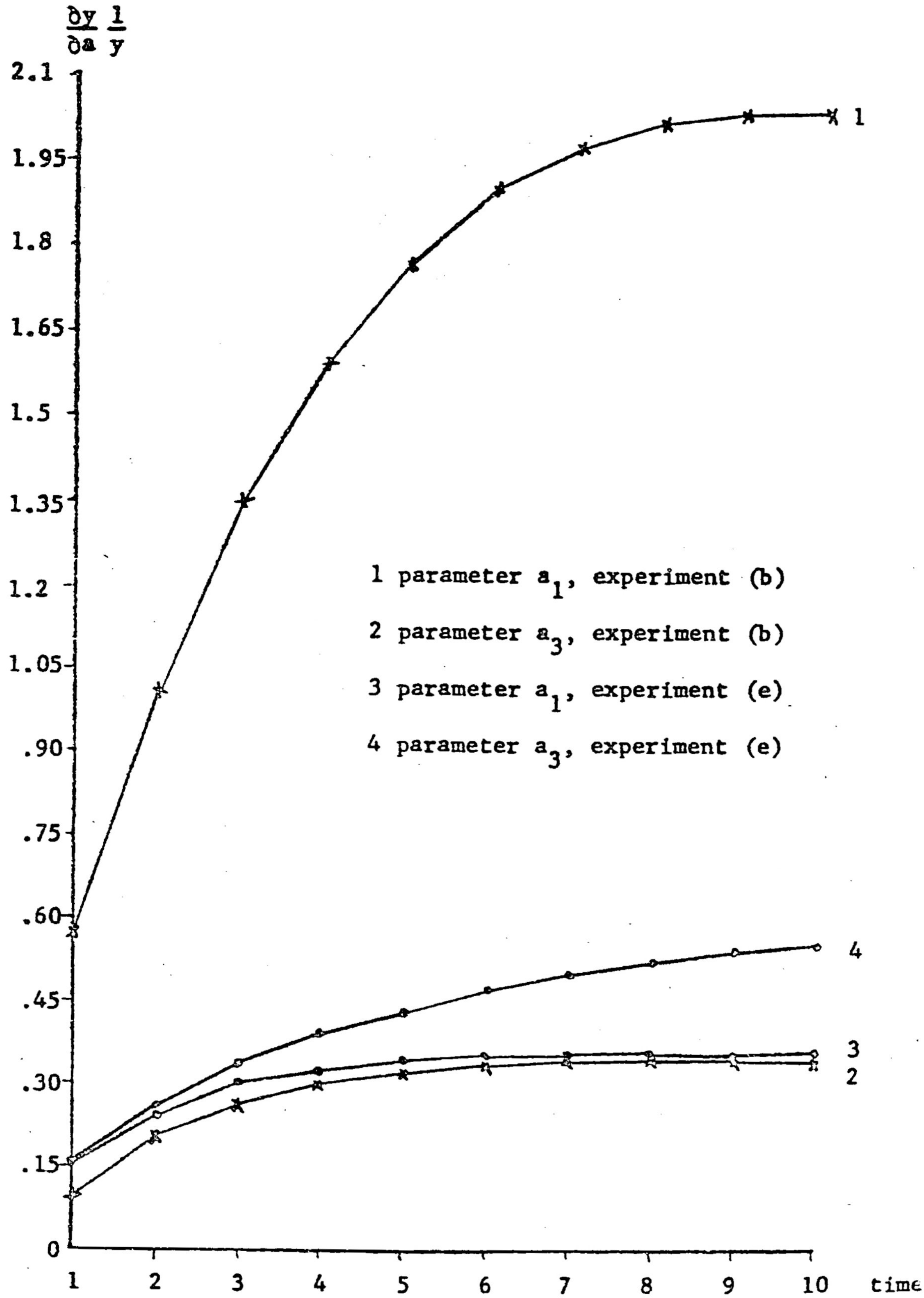

Figure (2) Normalized sensitivity coefficients associated with target variables.

tion in a_1. The reason, cf. equation (25), is that a change in a_1 will affect the corresponding state variable (X_1) in the same direction, but the net effect on output varibale will be very small since (X_1) appears in both numerator and denominator of the output variable.

Figure 3 shows the first column of matrix B in experiments (b) and (c) (only the effect on the second policy variable is plotted plotted). Again a_1 has more effect on the policy variable than a_4. This figure also shows that both effects decline over time (i.e., later values of policy variables were affected less than the earlier values). The reason for this will be explained shortly.

d) In all the experiments the rank of B (and hence of BVB') is two (the number of error terms), so U is restricted to just two dimensions in a linear subspace passing through $\bar{\underline{u}}^*$ (the center of the set, i.e., the nominal optimum). Thus the inequality:

$$\Delta \underline{u}' V_1^{-1} \Delta \underline{u} \leq 1$$

must be interpreted as first restricting $\underline{u}$ to this linear subspace and then to an ellipsoidal set U. By examining the axes of U, one can avoid using the inverse of V_1 explicitly; the axes are the basis for U. As an illustration, fig. 4-7 show the relative patterns of variation (the components of the eigenvectors of the matrix V_1 of unit length) in set U in experiments (b) and (c). The main interest here is in comparing the change in one policy variable relative to the change in the other. The 20 components of each eigenvector correspond to the 20 policy variables (two for each one of the ten planning years) defined in the experiments. By construction, the odd components of each eigenvector correspond to the first policy variable, while the even components correspond to the second policy

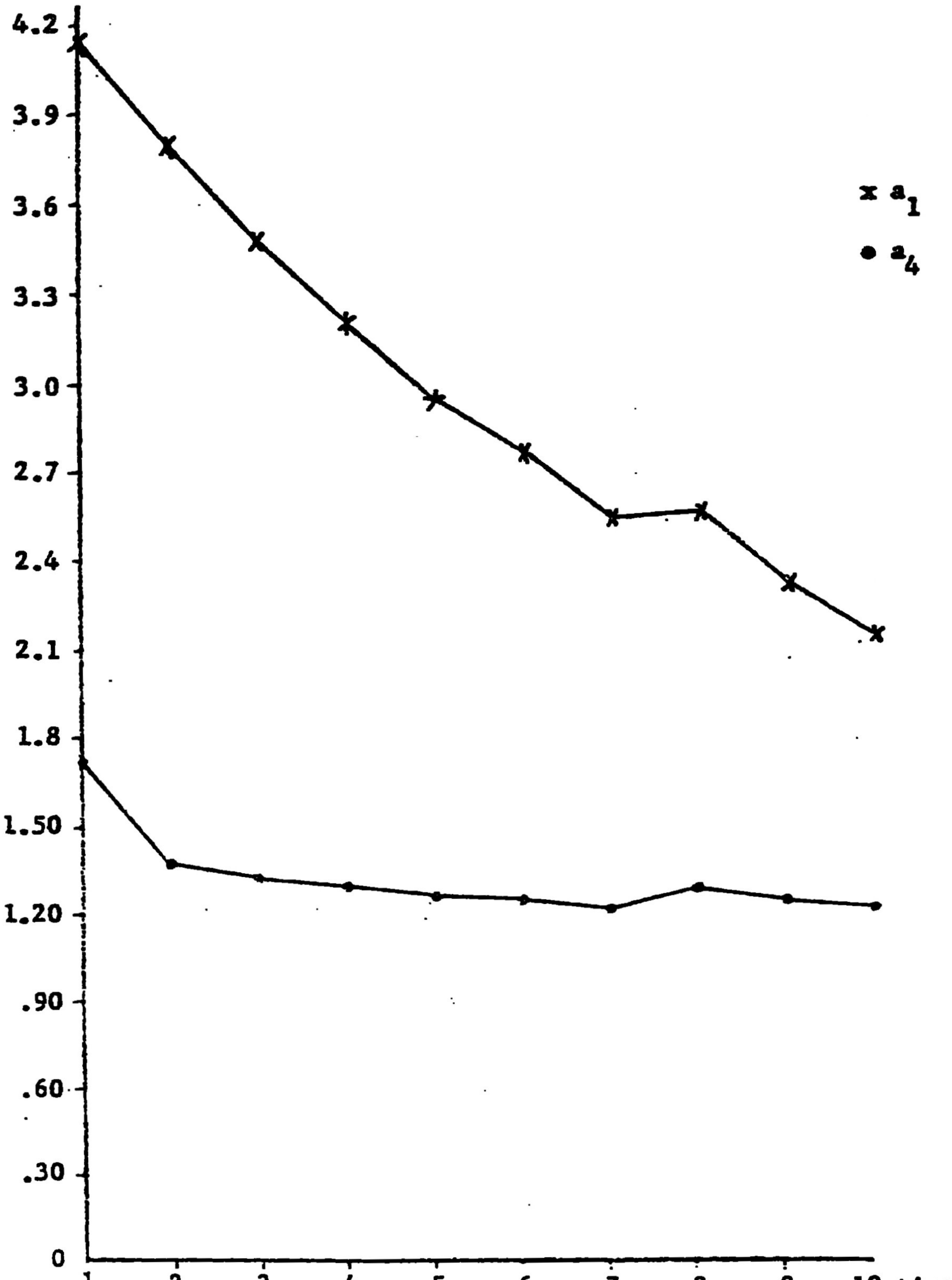

Figure (3) The shift in the optimum value of the second policy variable due to a unit change in the parameters a_1 and a_4.

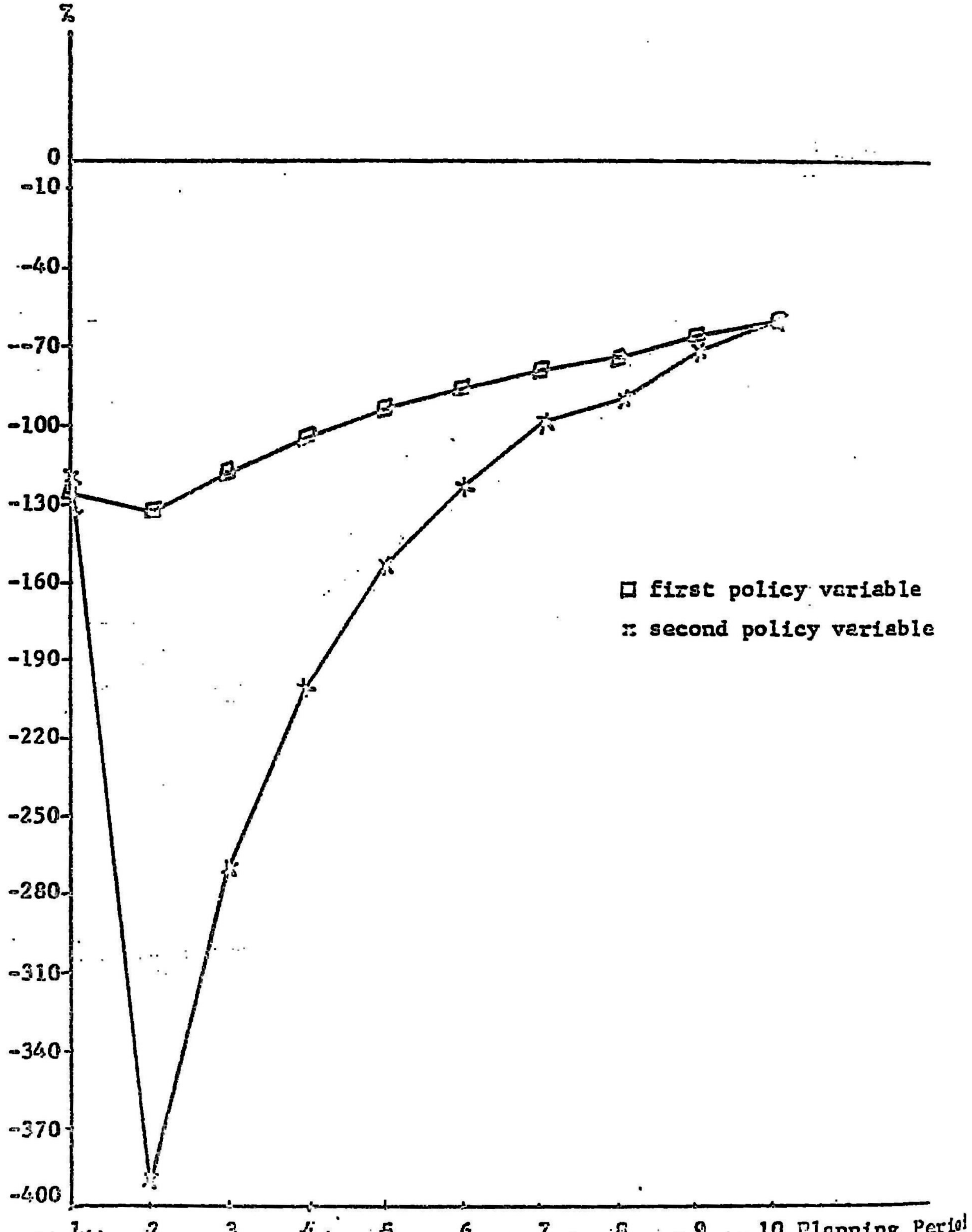

Figure (4) Components of the first eigenvector of unit length as a pattern of variations (percentage deviation of the nominal optimal value) in policy variables in experiment (b).

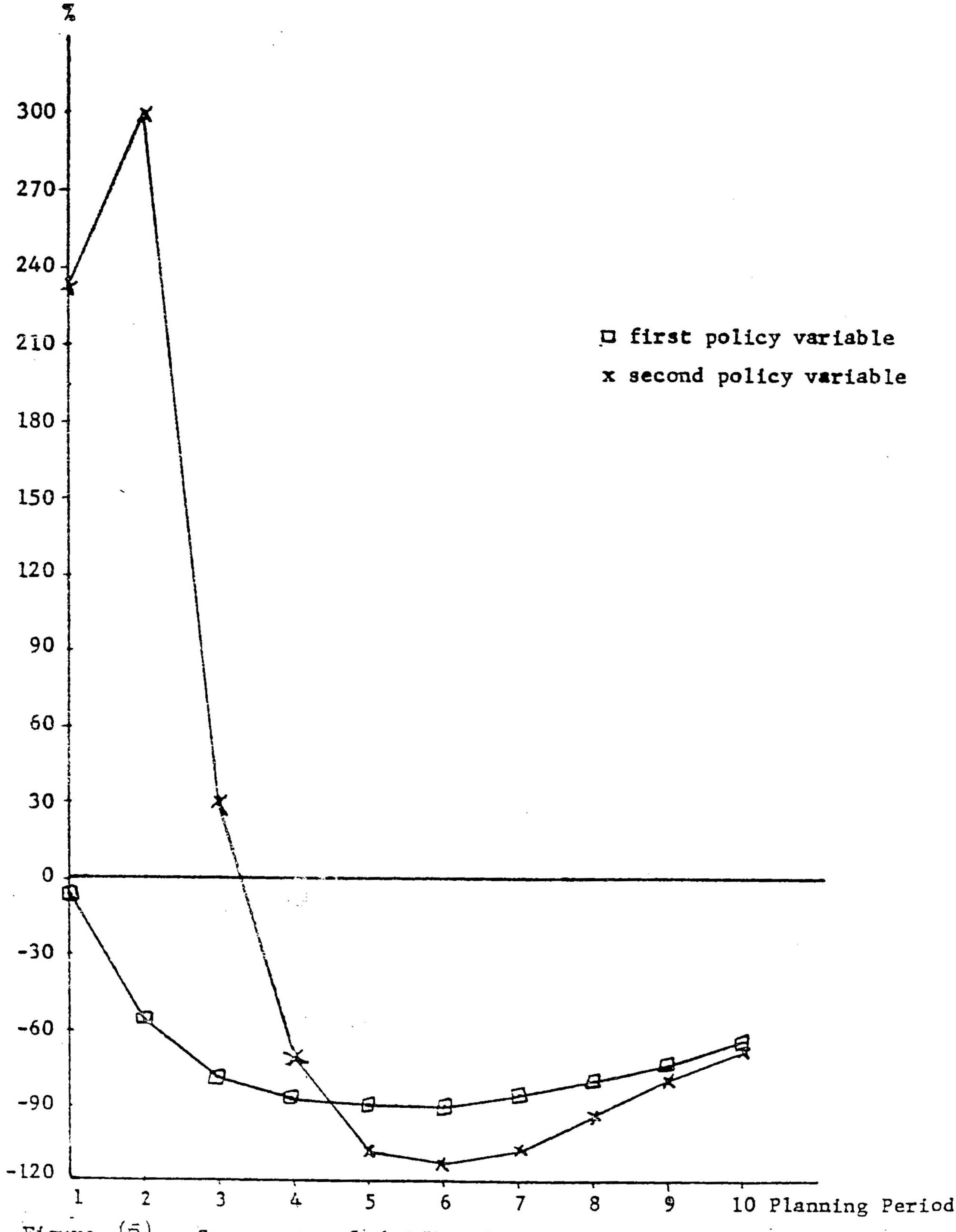

Figure (5) Components of the second eigenvector of unit length as a pattern of variations (percentage deviation of the nominal optimal value) in policy variables in experiment (b).

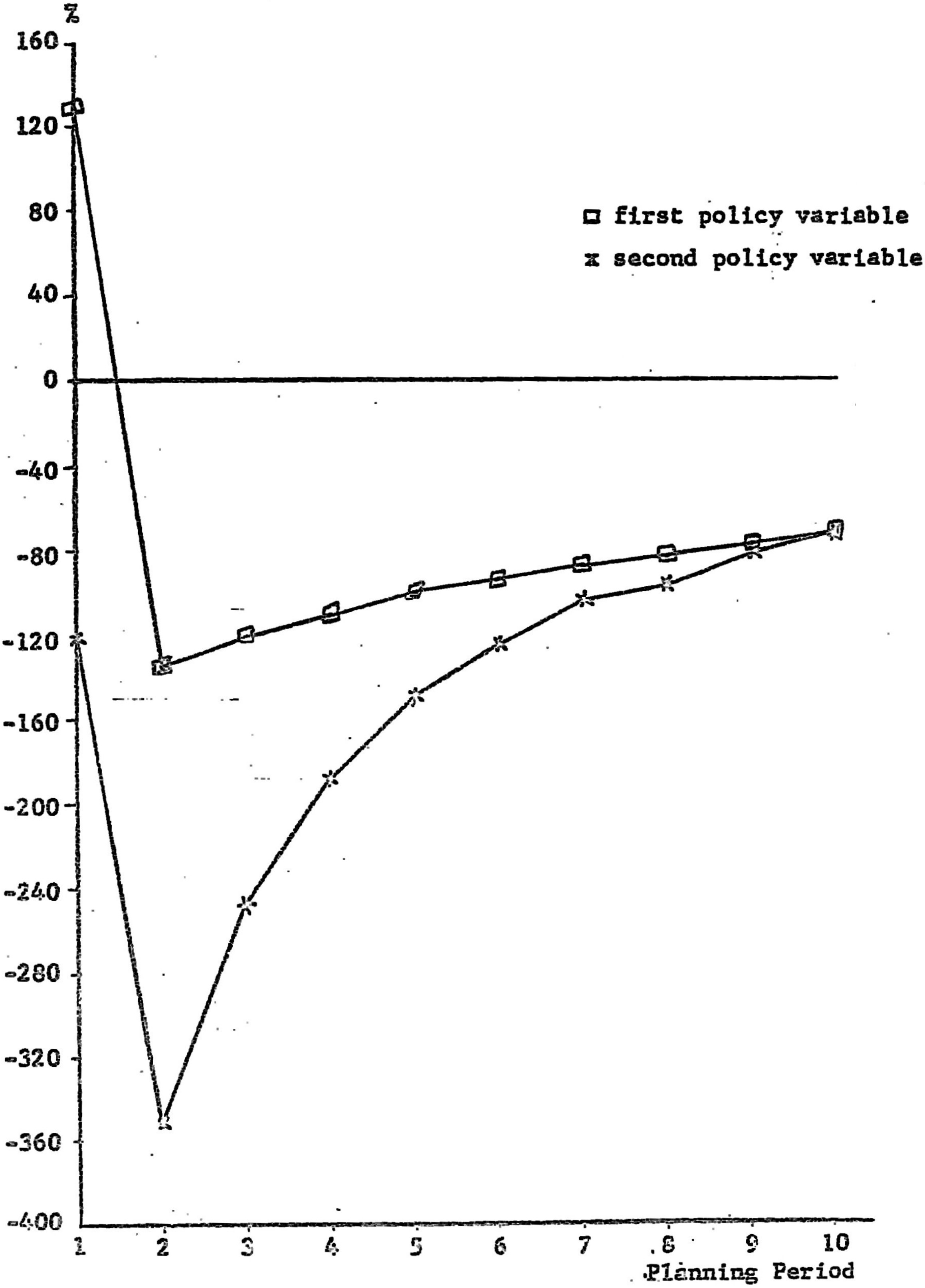

Figure (6) Components of the first eigenvector of unit length as a pattern of variations (percentage deviation of the nominal optimal value) in policy variables in experiment (c).

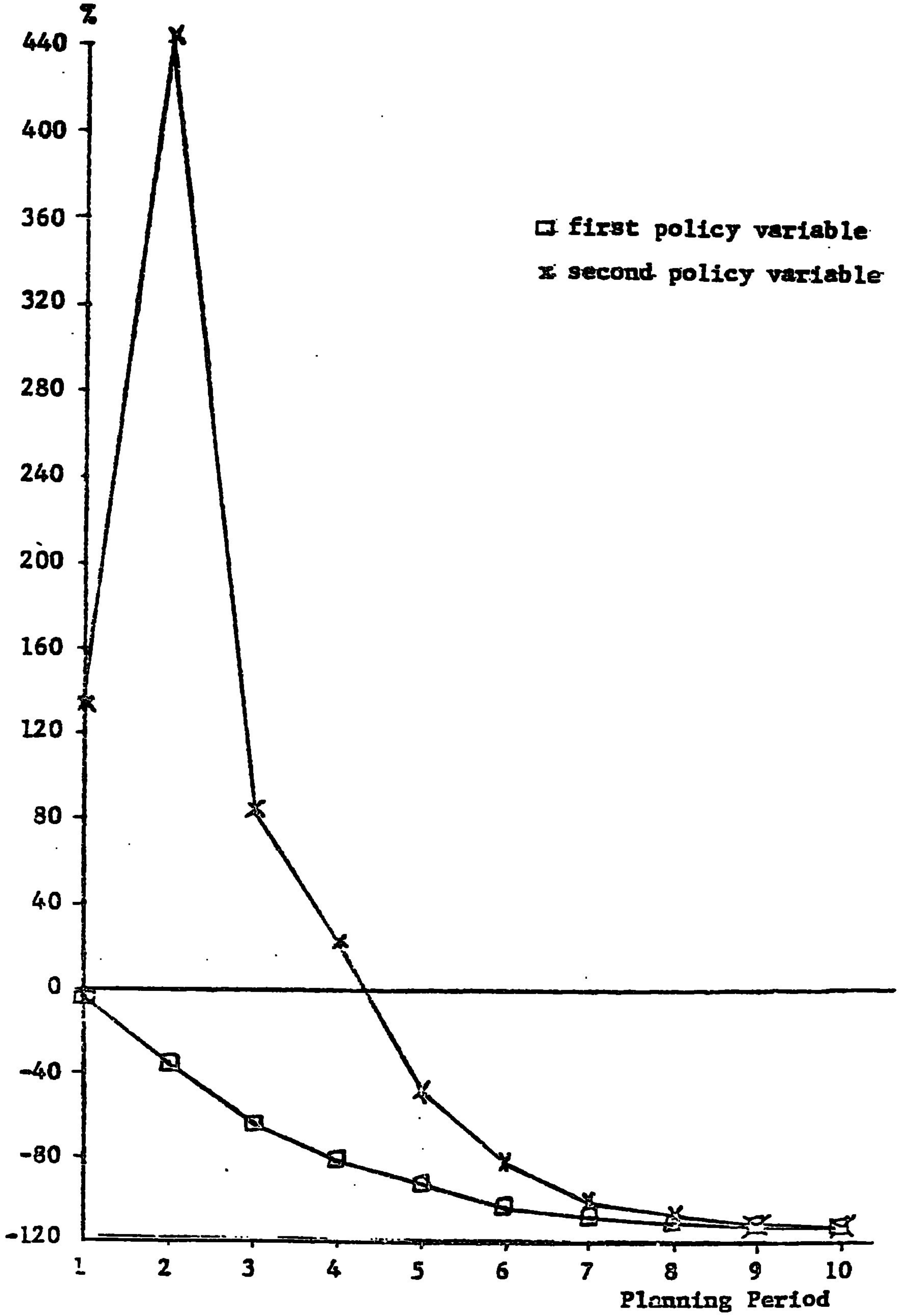

Figure (7) Components of the second eigenvector of unit length as a pattern of variations. (percentage deviation of the nominal optimal value) in policy variables in experiment (c).

variable. These two groups of components are plotted as two seperate curves as shown in these figures. Note that the matrix V_1 depends on the matrix V which depends in turn on the range of uncertainty assumed for each error term (denoted by $\pm e$ in section (3)) and which surrounds its nominal value $\bar{\epsilon}$.

Examining these figures, we observe that setU contains larger variations in the second policy varibale relative to the optimum but much smaller variation in the first policy variable. In all experiments later values of both policy variables were affected less than earlier values. This can be seen from the movement of both curves towards line (0) with variaiton in the second policy variable declining by a faster rate than that of the first policy variable. The explanation of these patterns of variaitons can be traced to the way these two policy variables affect the putput variables over time as exhibited in the matrix S_1 given by equation (27).

Table (1) shows the elements of this matrix in the first five periods only. The first observation about this matrix is that the sums of its columns are declining over time (i.e., for the later values of policy variable) which means that the "total" effect of policy variables on target variables is declining. This explains why the overall trend in the effect of errors on policy variables is also declining. The fact that the second policy variable is much more affected by deviations in error terms than the first due to the observation that although the column totals under both policy variables are declining, the total effect of the second policy variable is larger than that of the first. A final observation from Table 1 is that the effect of the first policy variable on target variables although less in total than that of the second policy variable, is

Table (1)

The Effect of a Unit Deviation in Policy Variables
(in the First 5 Periods) on the Target Variables
Experiments (a), (b) and (c)

Period	1		2		3		4		5	
Policy Variable	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
	65	338	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	113	331	62	362	0.0	0.0	0.0	0.0	0.0	0.0
	161	325	116	355	68	388	0.0	0.0	0.0	0.0
	209	318	168	348	130	380	75	415	0.0	0.0
	260	313	223	342	188	373	144	409	82	446
	305	306	274	335	246	365	205	401	158	437
	363	300	336	331	308	360	281	395	233	428
	425	292	390	318	370	350	342	382	301	418
	486	286	466	311	445	344	418	374	384	411
	562	281	534	311	520	339	493	370	459	404
Total	2949	3090	2569	3013	2275	2899	1958	2746	1617	2544

increasing for later target variables while the reverse is true for the second policy variable. This may explain why the effect of deviations in error terms on the second policy variable is declining at a much faster rate than in the case of first policy variable.

To comapre the joint effect of different groups of errors among different experiments, Figures 8 and 9 show the components of the longest axis (i.e., the normalized eigenvector multiplied by the square root of the corresponding eigenvalue) of set U in experiments (a), (b) and (c) for the two policy variables. These experiments differ only in what constitutes the error terms. These figures reveal the same general characteristics discussed previously and need no further discussion except to note that these two figures reflect only the differences in the individual effect of the error terms on target varibales which are represented by the matrix S_{ϵ} .

e) The case of interacting deviaitons. In all of the discussion so far we have assumed that although the deviations in the error terms have to satisfy the equation that describes the elliptical set E (which puts a restriction on the magnitude of these deviations), they are allowed to move independently. This was implicitly assumed when we defined the matrix of the quadratic form of set E, the matrix V, to be a diagonal matrix.

When the off diagonal elements of the matrix V are nonzero (we will assume without loss of generality that these elements are equal i.e., the matrix V is symmetrical), the ellipse will be in a titled position with the major axis being either sloped negatively when the off diagonal elements are negative, or sploped positively when the off diagonal elements are positive. In the first case (negative slope), the deviations in the error terms would move in the opposite

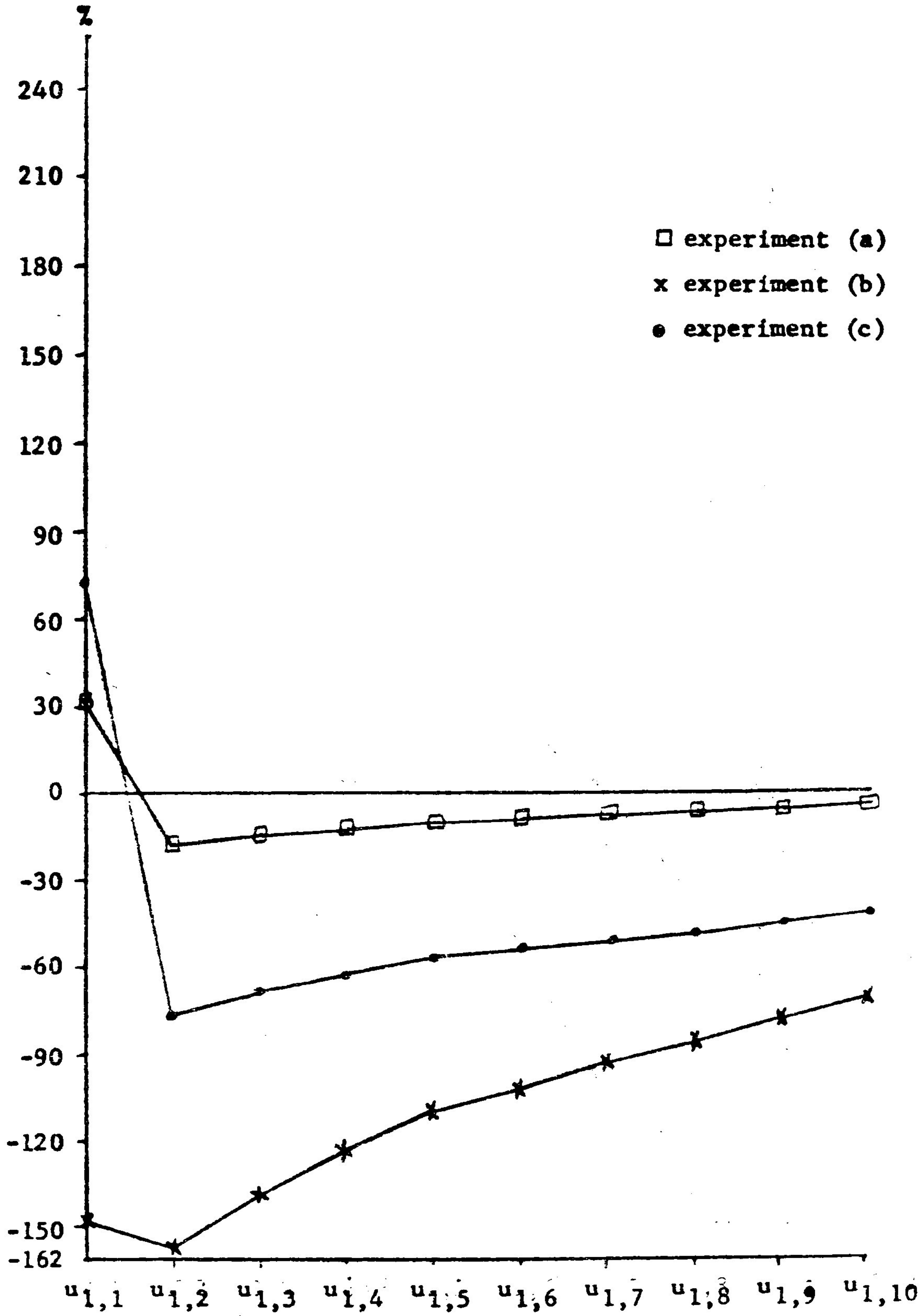

Figure (8) Component of the longest axis as a pattern of variation (percentage deviation of the nominal optimal value) in the first policy variable in experiments (a), (b) and (c)

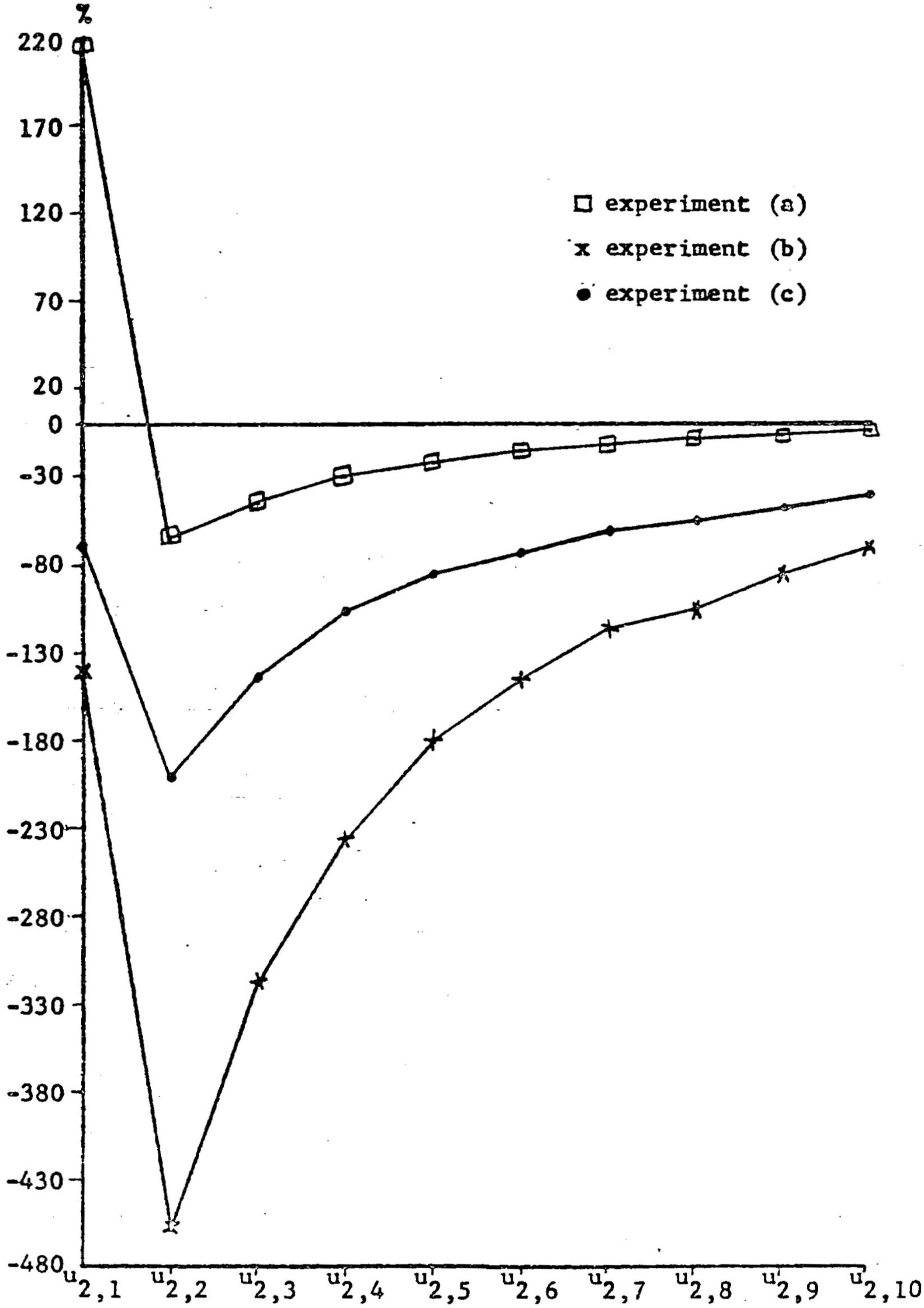

Figure (9) Component of the longest axis as a pattern of variation (percentage deviation of the nominal optimal value) in the second policy variable in experiments (a), (b) and (c) directions, i.e., the error term would tend to err in the opposite directions. This a positive deviation in the first error term, for example ($\epsilon_1 > \bar{\epsilon}_1$), will be accompained by a negative deviation in the second error term ($\epsilon_2 < \bar{\epsilon}_2$) and vice versa. This case will be referred to as the case of negative interaction. In this second case (positive slope), the error terms would tend to err in the same direction, i.e., a positive (negative) deviation in the first error term will be accompained by a positive (negative) deviation in the second. This case will be called the case of positive interaction. To distinguish the case which we have been discussing so far, we will refer to it as the case of zero interaction.

That the deviations can be expected to move in a known direction (same or opposite) may well be the case in certain situations. Fo example, when a group of exogenous variables or parameters are projected or estimated by the same method, it is not unplausible to assume that over-estimating one parameter is most likely to be accompanied by over-estimating the second parameter or vice versa.

To examine the effect of these concepts, the two cases of positive and negative interaction were assumed for the deviations in error terms in experiments (b) and (e).[10] The off diagonal elements assumed in the experiments were such that (together with diagonal elements) they are equivalent to a correlation coefficient of .60 in experiment (b) and .70 in experiment (e).

Figures 10 and 11 show the components of the longest axis as a pattern of variation in the two cases of positive and negative interaction in these two experiments. Table 2 shows the results for some selctive measures for the two cases of positive and negative interaction together with the case of zero interaction for comparison.

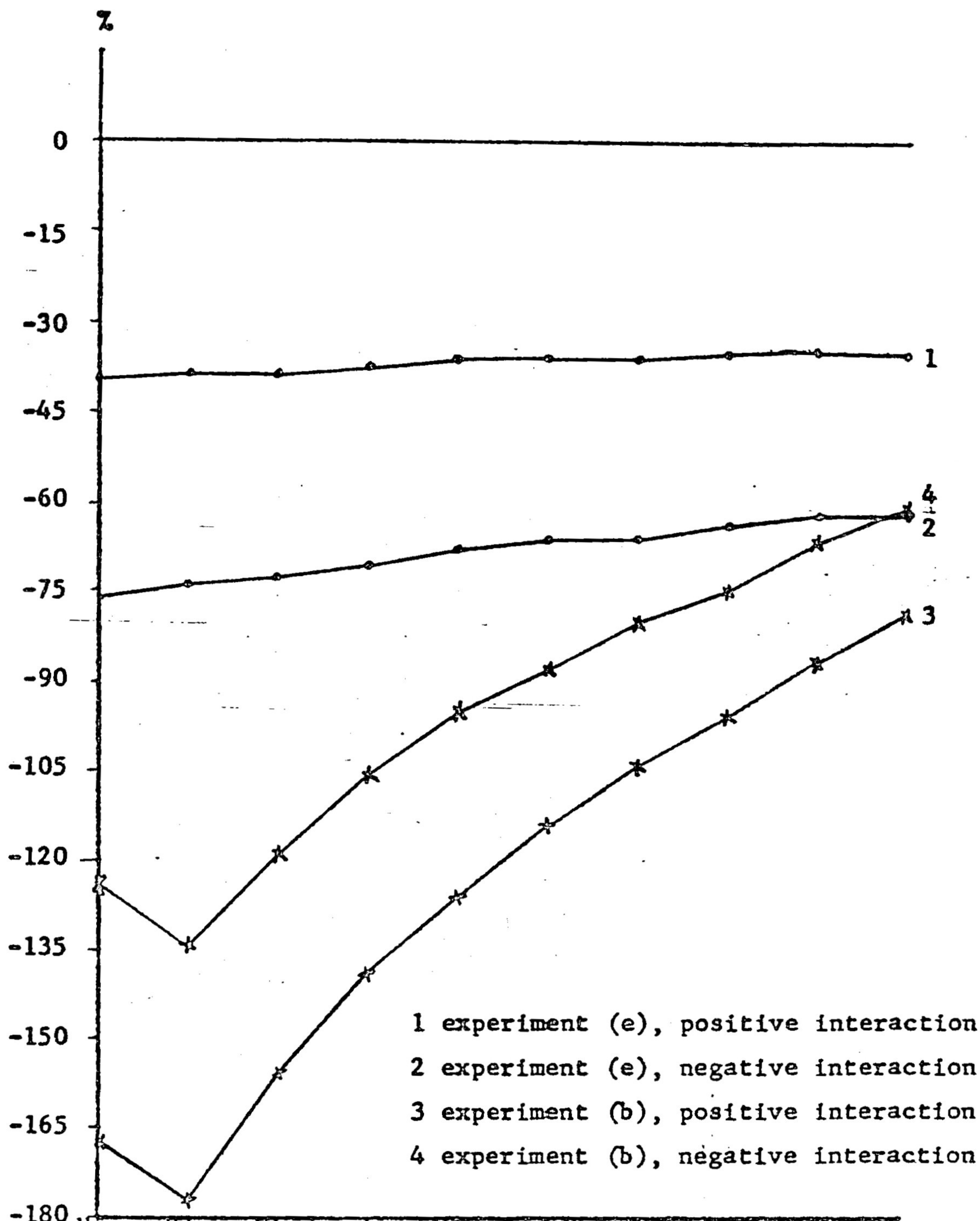

Figure (10) Component of the longest axis as a pattern of variation in the first policy variable (percentage deviation of the nominal optimal values) in the case of positive and negative interaction between the error variable; experiments (b) and (e)

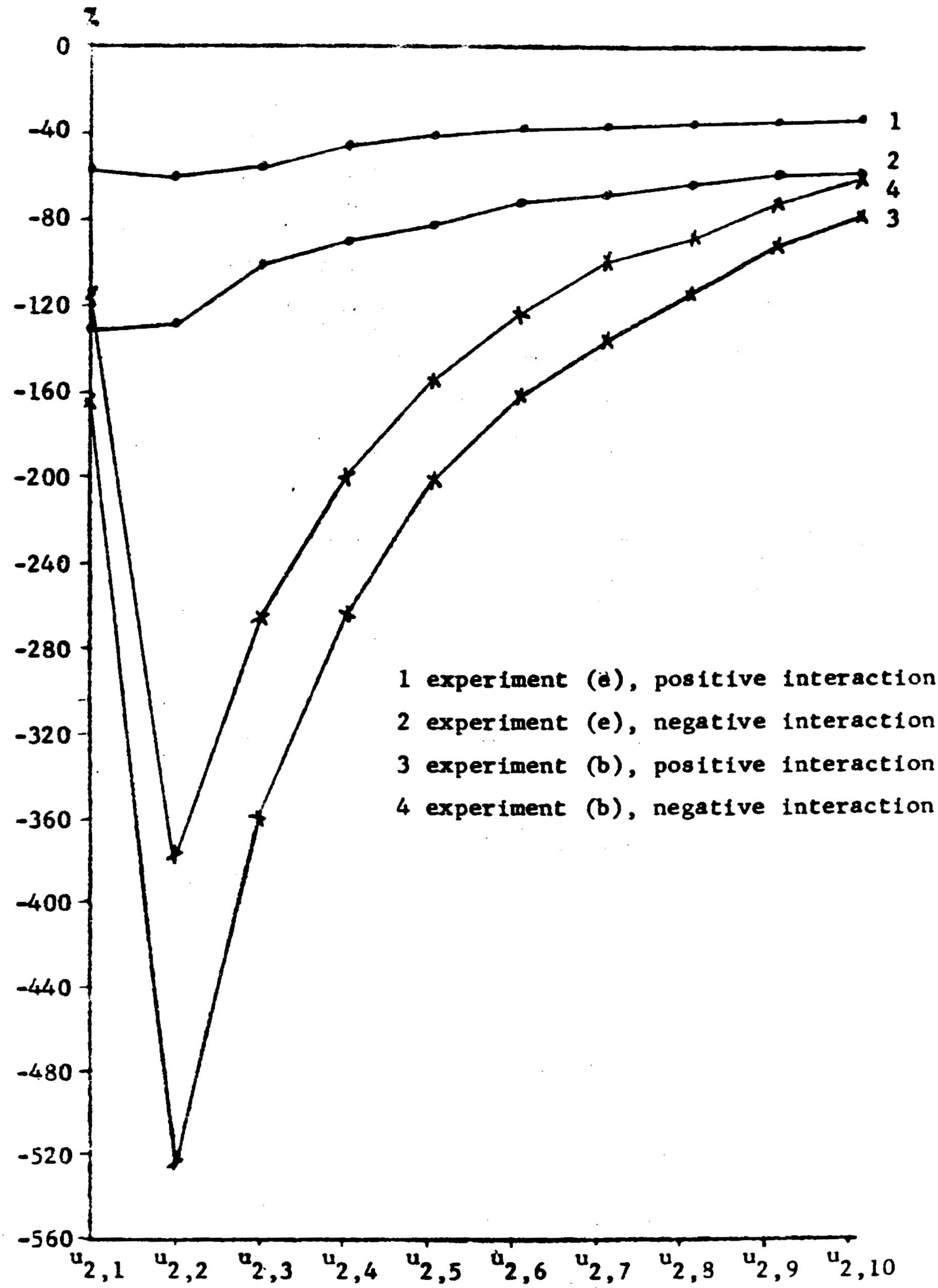

Figure (11) Component of the longest axis as a pattern of variation in the second policy variable (percentage deviation of the nominal optimal values) in the case of positive and negative interaction between the error variable; experiments (b) and (e)

Table (2)

Some Measures of Variation in the Set U, in the Case of Zero, Positive and Negative Interaction between the Error Terms Experiment (b): $\epsilon_1 = u_1$, $\epsilon_2 = a_3$

Interaction	Positive		Zero		Negative	
Coefficient of Variation of $(\sigma/\bar{u}^*) \times 100$	1st Policy Variable	2nd Policy Variable	1st Policy Variable	2nd Policy Variable	1st Policy Variable	2nd Policy Variable
	35.70	34.93	31.48	30.15	26.58	24.46
	37.74	111.99	33.43	97.69	28.49	80.89
	33.35	76.82	29.65	67.57	25.41	56.84
	29.56	56.65	26.34	50.15	22.68	42.66
	26.24	43.28	23.42	38.50	20.22	33.03
	24.31	34.72	21.73	30.97	18.79	26.67
	22.18	27.74	19.84	24.80	17.18	21.47
	20.61	25.04	18.43	22.38	15.96	19.36
	18.49	20.14	16.55	18.02	14.34	15.61
	16.81	16.81	15.04	15.04	13.04	13.04

	Positive	Zero	Negative
Length of longest axis	l_1 = 1.3296	1.1712	.9883
Length of second axis	l_2 = .0281	.0394	.0378
l_1 / l_2	47.4	29.8	26.2

Table (2) Cont.

Experiment (e) : $\epsilon_1 = a_1$, $\epsilon_2 = a_3$

Interaction	Positive		Zero		Negative	
Coefficient of Variation of ($\sigma/\bar{u}$) x 100	1st Policy Variable	2nd Policy Variable	1st Policy Variable	2nd Policy Variable	1st Policy Variable	2nd Policy Variable
	8.51	12.46	12.98	21.43	16.26	27.63
	8.42	12.94	12.76	21.43	15.96	27.40
	8.24	10.46	12.40	16.86	15.49	21.43
	8.07	9.74	12.07	15.30	15.04	19.32
	7.82	9.16	11.61	14.11	14.44	17.72
	7.68	8.23	11.33	12.47	14.06	15.59
	7.73	8.04	11.33	11.98	14.04	14.92
	7.58	7.64	11.05	11.24	13.67	13.94
	7.39	7.21	10.70	10.49	13.20	12.97
	7.50	7.10	10.80	10.22	13.30	11.13

	Positive	Zero	Negative
Length of longest axis	ℓ_1 = .0789	.1205	.1510
Length second axis	ℓ_2 = .100535	.00506	.0028
ℓ_1/ℓ_2	14.75	23.8	53.95

In this table the coefficient of variation was obtained using the assumption of uniform distribution in set U and Eq. (21) and using the nominal optimum vector $(\underline{\bar{u}}^{*})$ as the mean vector.

In Figures 10 and 11 the higher value assumed for the effect of interacted deviations in experiment (e) resulted as expected in a larger difference between the two cases than in experiment (b). An interesting observation is that in experiment (e) the larger variation resulted when there was a negative interaction between the deviations in the error terms and not with a positive interaction as in the case of the other experiment. This observation is also seen in Table 2 whatever the measures of comparison were, coefficient of variation or length of axis (the longer the axis, the the variations). Before commenting on the last observation, let us examine Table 2 more closely.

1) There is a very large difference in the length of the two axes. This means that practically all variations are concentrated (or moving) in one direction. This is clear from all the figures in all experiments where variations in the two policy variables move together in the same direction (towards zero).

2) Examining differences between the three cases of interaction in experiments (b) and (e) shows that the larger the variation (positive interaction in (b) and negative interaction in (e) in terms of the coefficient of variation, the larger the difference between the length of the two axes (or vice versa). This may be due to the individual effect of deviations in error terms on policy variables as reflected in matrix B. Examining the columns of matrix B in all experiments (not shown here) together with the

the results of zero interaction in this table, we observe a general trend ; the larger the differences between the individual effects of error terms on policy variables, the larger the differences between the length of the two axes. This shows how the individual effects of the deviations interact with their amplitudes to shape set U.

3) The largest variations in experiment (b) occurred with a positive interaction, while the largest variations in experiment (e) occurred with a negative interaction. Recalling the basic difference between experiments (b) and (e) (namely that in experiment (b) changes in error terms affect target variables in the same direction as their effect on state variables, while in experiment (e) the two groups of changes move in the opposite direction), we arrive at the following conclusion: the larger variation in policy variables occurs when the relationship between the effect of deviations in error terms on state variables and the effect of changes in state variables on target variables is the same as the relationship between those deviations.

Finally, from the experience I had in applying the above methodology to the example model, some remarks which may be helpful in applying it to larger scale simulation models can be made.

1) Educational simulation models tend to be large because of more equations and not because of more variables per equation. In the approach presented in chapter three, the simulation model used to provide the value of the partial derivatives needed for the optimization process. These partial derivatives were computed numerically using the finite difference

approach. Most of the standard optimization algorithms that use these numerically computed derivatives spend most of their time doing function evaluations. One function evaluation in the present context corresponds to the simulation of a $r+m_1+m_2$ - equation model for N periods (here R is the number of state variables in the model, m_1 is the number of output variables, m_2 is the number of policy variables, and N is the planning interval) plus the rather trival computation of the objective function. Thus, the cost for solving the optimization problem depends mainly on the number of the functions evaluated. In some of the experiments discussed in the text, it took about 5.23 seconds, 9 iterations and 100 function evaluations to find the optimum solution.

2) One way to save the time needed for function evaluations is to write the program that does these evaluations (i.e., that simulates the whole model) in such a way that no computations are performed other than those that are absolutely needed to go from the initial values of the policy variables (needed to begin the simulation) to the value of the objective function (at the end of the simulation). For example, any set of calculations using exogenous variables that is not changed as a result of changes in the value of the policy variables should not be done in the function-evaluation program. Rather, these calculations should be done only once before the solution of the optimal policy problem begins. This procedure was followed to the best of my ability in the computations done in this paper.

3) Another way to cut down the time needed to solve the optimization problem is to choose a good starting point for the

policy variables. A natural starting point is their desired values which appear in the objective functions. This was found to increase the speed of concergence in some of the experiments which resulted in cutting down the number of function evaluations by 10%, the number of interations required by 23%, and the time by 23%. A better starting point, which may result in a significant saving of time, can be obtained by first solving a small problem and then using the answer to this problem as a starting point for the larger problem. I have used the optimum solution to a smaller but otherwise equivalent problem (the planning period was 5 years instead of 10) as the starting point for the original problem. That resulted in cutting down the time substantially (the number of function evaluations was decreased by 35%, the number of iterations by 45%, and the time by 37%).

4) In terms of the size of the problems of the type treated here that the standard optimization algorithms can handle, there is an obvious trade-off between the size of the model, the number of policy variables, and the length of the planning period. Unfortunately, without actual experimenting with different large scale educational flow models, it is hard to establish any accurate values to determine what problems are practical to solve and what are not. On the other hand, the major advantage of the procedure introduced in this paper is that it builds on the computations that are usually carried out during the actual planning stages and that was discussed in the introduction. Thus, the simulation experiments usually done using education flow models, together with the computer program that are used to perform those experiments, are the basic building blocks

in the procedure presented here. The additional computational steps needed to yield the measures aimed at examining the effect of uncertainty can easily be added to and incorporated into the existing steps to provide a homogenous and effective package.

References

1. Armitage, P., C. Smith and P. Alper, Decision Models for Educational Planning. London: Penguin Books, 1969.

2. Cramer, H., Mathematical Methods of Statistics. Princeton: Princeton University Press, 1964.

3. Davis, R. G. and E. Schiefelbein, Development of Educational Planning Model and Application in the Chilean School Reform. Lexington, MA.: Lexington Books, 1974.

4. Hooker, B. L., Educational Flow Models with Application to Arab Statistical Data. Sweden: Linkoping School of Education, 1974.

5. Murray, E. (ed.), Numerical Methods for Unconstrained Optimization. New York: Academic Press, 1972.

6. OECD, Mathematical Models in Educational Planning. Paris: OECD, 1967.

7. OECD/CERI, S.O.M. : A Simulation Model of the Educational System. Technical Report. Paris: OECD, 1970.

8. Stone, R., "A Model of the Educational System." Minerva, Winter, 1965.

9. ——————, Input-Output and Demographic Accounting: A Tool for Educational Planning. MIH, 1966, pp. 365-381.

10. ——————," A View of the Conference." In OECD, Mathematical Models in Educational Planning. Paris: OECD, 1977, pp. 7-20.

11. UNESCO, An Asian Model of Educational Development: Perspectives for 1965-80. ED. 6610.33/4, 1966.

12. Werdelin, I., Quantitative Mehtods and Techniques of Educational Planning. Beirut: Regional Centre for Educational Planning and Administration in the Arab Countries, 1972.

13. Zelinsky, O. , A First Course in Linear Algebra. New York: Academic Press, 1973.

Footnotes

1. It is important to note that the amount of deviation given by $\delta \underline{\epsilon} = \underline{\epsilon} - \bar{\underline{\epsilon}}$ is not a random quantity that can be randomly during the optimization process. But rather, once it is determined at the beginning of the iterations, it remains fixed throughout the procedure. Thus, our conceptualization of the uncertainty here rests on the observation that the amount of deviation $\delta \underline{\epsilon}$, although it remains fixed once determined, represents only one possible realization of many alternative values.

2. Let $n_3 = 2$ and let f (ϵ_1, ϵ_2) be the joint density function defined on the set (ellipse)E. Then, if (f) is defined such that:

$$f(\epsilon_1, \epsilon_2) = \text{Prob.}((\epsilon_1, \epsilon_2) \text{ in } E) = \begin{cases} \frac{1}{M} & (\epsilon_1, \epsilon_2) \text{ in } E \\ 0 & \text{otherwise} \end{cases}$$

where M is the area of the ellipse E, and are said to be uniformly distributed on E. The range of uncertainty (+e) has to be relatively small for the uniformity assumption to be consistent with the restriction on the magnitude of the parameters of the model.

3. See Cramer (1946) pp. 284-285 and pp. 300-301.

4. See Zelinsky (1973) pp. 217-219 and Ch. 5.

5. This is in accordance with optimization procedure used in computation when the objective function consists of sum of squares.

6. A retention ratio of .98 was used to compute the number of teachers still available in the system each year.

7. Those weights reflected the differences in units of measurement between the target and policy variables in addition to giving more priority to keeping the first policy variable near its desired path.

8. This was the case in experiments (a), (b), (c) and (d).

9. This was the case in experiment (e).

10. Since only the matrix V will be affected by these new assumptions, only the second group of measures will reflect the changes.

Acknowledgment

This paper has been extracted, with some modifications, from the author's doctoral dissertation completed at the Johns Hopkins University, U.S.A.

The author very gratefully acknowledge his intellectual debt to faculty members at the Department of Population Dynamics and especially to Professor I. Sirageldin, the thesis advisor, for a rich and stimulating experience. Very special thanks go to the Population Council for New York,U.S.A. for the financial support given to the author during his graduate studies.

---

* Assistant Professor, Department of Biostatistics and Demography Institute of Statistical Studies and Research, Cairo University.