This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Annals of Economic and Social Measurement, Volume 1, number 4

Volume Author/Editor: NBER

Volume Publisher:

Volume URL: http://www.nber.org/books/aesm72-4

Publication Date: October 1972

Chapter Title: Linear Decision with Experimentation

Chapter Author: Elizabeth Chase MacRae

Chapter URL: http://www.nber.org/chapters/c9446

Chapter pages in book: (p. 435 - 445)

*Annals of Economic and Social Measurement*, 1/4, 1972

# LINEAR DECISION WITH EXPERIMENTATION

BY ELIZABETH CHASE MACRAE*

*A solution to a structural approximation of the adaptive linear decision problem with unknown parameters is derived and given an economic interpretation in terms of the price and stock of information. Numerical implementation of the solution indicates that when learning is taken into account, the optimal policy paradoxically can employ less experimentation and, thus, cause less to be learned than when learning is not taken into account in the derivation of the solution.*

## 1. INTRODUCTION

A common procedure for dealing with economic decision making models is first to estimate the parameters of the relationships between the policy and endogenous variables, and then to carry out policy action based on those estimates. This technique can be broadened somewhat to incorporate variances of the estimates into the problem as a measure of uncertainty but the basic separation between estimation and control remains. Work along these lines has been carried out in the context of a linear decision model by Holt, Modigliani and Muth [3] and by Theil [9] for the certainty equivalence case, by Brainard [2] for a simple static model with uncertainty, and by MacRae and MacRae [7] for a general multivariate dynamic model.

However, if the unknown parameters of the model are assumed to remain constant over time, an optimal decision can take advantage of future observations as they are received to learn more about the parameters. In addition, the choice of values for current policy variables can affect how much is learned next period. In other words, current policy decisions have a twofold function in a model with unknown parameters: they can directly control the endogenous variables, albeit with some error, and they can be used to learn more about the unknown parameters so that future control errors may be reduced. There is no inherent value in learning for its own sake in the decision making problem. The only value comes from improved future control, and the only costs are those incurred by sacrificing current control for experimentation.

The purpose of this paper is to analyze the interaction between estimation and control in the context of a scalar linear decision model with unknown parameters.[1] The unknown parameters are modelled as Bayesian random variables with means and variances which change as additional observations are received over time. Since the mathematical problem is not analytically soluble, an approximation is presented which preserves the nature of the interaction between estimation and control, but which allows a solution to be derived. The equations characterizing the solution can be interpreted in terms of such economic concepts as price of

* Part of the work on this paper was done while the author was with the Division of Research and Statistics, Board of Governors of the Federal Reserve System.

[1] For an analysis of the multivariate model see MacRae [6].

information and value of estimating, and provide some insight into the structure of an optimal solution. Finally, several numerical examples are given.

## 2. Problem

The decision maker is assumed to be faced with the problem of choosing values for policy or control variables for $N$ periods so as to minimize the expected value of a quadratic objective function subject to a stochastic linear difference equation of the form

$$x_{j+1} = ax_j + bu_j + c + \varepsilon_j, \quad j = 0, \ldots, N-1 \tag{2.1}$$

where $x_j$ and $u_j$ are the values of the endogenous and policy variables respectively in period $j$, and $\varepsilon_j$ is a normally distributed noise term with zero mean, known contemporaneous variance $\Omega$, and zero intertemporal covariances. The parameters $a$, $b$ and $c$ are unknown but are assumed to be constant over time. To capture the uncertainty regarding the values of $a$, $b$ and $c$ they are modelled as Bayesian random variables with conditional means $a_j$, $b_j$ and $c_j$ and conditional covariance matrix $\Gamma_j$, where the subscripts indicate values based on all observations through $x_j$ and $u_{j-1}$.

To simplify the notation, equation (2.1) may be rewritten as

$$x_{j+1} = d'w_j + e_j \quad j = 0, \ldots, N-1, \tag{2.2}$$

where $d$ and $w_j$ are vectors defined by

$$d' = (a, b, c) \tag{2.3}$$

and

$$w_j' = (x_j, u_j, 1) \tag{2.4}$$

If a normal prior is assumed for $d$ (now considered a random vector) with mean $d_0$ and variance $\Gamma_0$ then subsequent posterior or conditional distributions will also be normal, and it can easily be shown that the conditional means and variances can be computed recursively by

$$\Gamma_{j+1}^{-1} = \Gamma_j^{-1} + w_j w_j'/\Omega \tag{2.5}$$

and

$$d_{j+1} = \Gamma_{j+1}(\Gamma_j^{-1} d_j + w_j x_{j+1}/\Omega). \tag{2.6}$$

The inverses of $\Gamma$ in (2.5) and (2.6) are understood to be pseudoinverses if $\Gamma$ is singular because $a$, $b$ or $c$ are known with certainty.

The decision maker chooses his control variables over time so as to minimize the expected value of a quadratic objective function:

$$J = E\left\{ \sum_{j=1}^{N} \tfrac{1}{2}q_j x_j^2 + \tfrac{1}{2}r_j u_{j-1}^2 + s_j x_j + t_j u_{j-1} | x_0 \right\}, \tag{2.7}$$

where $q_j$ and $r_j$ are assumed to be such that an optimum exists. It is further assumed that the decision maker will have observations available through $x_j$ when he is faced with the problem of optimally choosing a value for $u_j$, so that future policy variables may be specified in terms of the unknown future values of $x$. The expected value in (2.7) is taken with respect to $a$, $b$, and $c$ as well as $\varepsilon$.

## 3. METHOD

It is well known that the problem as stated in Section 2 is not analytically soluble in general.[2] This is because of the interaction between the random variables $a$, $b$ and $c$ and their means and variances, which are also random since from (2.5) and (2.6) they can be seen to depend upon future, unobserved values of $w$ and $x$. Because the problem cannot be solved directly, some sort of approximation must be used.

There are two general types of approximations which are generally used, which may be called numerical and structural. In a numerical approximation, the intractable functions are replaced by a finite set of points on the functions, calculated for a specific set of parameter values and initial conditions, while in a structural approximation the intractable functions are replaced by analytic functions which are similar in form. Although numerical approximations are often incorrectly thought of as producing a "true" optimal solution, it should be borne in mind that both types involve an alteration of part of the problem.[3] In this paper a structural approximation will be used to produce a framework which can be used to gain insights into the behavior of an optimal solution.

As was noted above, the mathematical difficulty in the problem arises from the fact that not only are the parameters, $a$, $b$ and $c$ random, but their future means and variances are also random. Previous work has dealt with this problem by constructing a sequence of open loop subproblems, starting in different periods, of which only the initial policy values are actually implemented. In each subproblem some of the randomness of the original is ignored so that a solution may be derived, but once the initial policy for that subproblem has been applied, complete updating of the parameter means and variances is carried out to provide the prior information for the next open-loop subproblem. Murphy [8], in each open loop, assumes that optimal future policy variables are linear functions of predicted endogenous variables, with the prior values for the parameters being used in the prediction. This assumption reduces the open loop problem to a deterministic problem with a changing but nonrandom parameter variance. Another quite different approach is given by Tse and Athans [10]. They update neither the means nor the variances of the parameters in each open loop, but incorporate an indirect learning feature by explicitly using the fact that the unknown parameters are constant over time, which affects the variance of future predictions for the endogenous variables.

In both of the above papers the open-loop problems are designed to produce deterministic values for the policy variables, based on predicted average behavior of the system. This paper preserves the basic stochastic nature of the open-loop

[2] See Aoki [1].

[3] The optimal solution calculated by Prescott elsewhere in this volume is an example of a numerical approximation.

problems by using predicted future variables only in the update equations for the means and variances. For the open-loop problem beginning in period 1 (with prior information available in period 0), the modified update rules become

$$\Gamma_{j+1}^{-1} = \Gamma_j^{-1} + E\{w_j w_j' | x_0\}/\Omega \tag{3.1}$$

and

$$d_{j+1} = \Gamma_{j+1}[\Gamma_j^{-1} d_j + E\{w_j x_{j+1} | x_0\}/\Omega]. \tag{3.2}$$

One implication of this approximation is that the updated value of $d_{j+1}$ is equal to the preceding value, $d_j$. For,

$$\begin{aligned} E\{w_j x_{j+1} | x_0\} &= E\{w_j E\{x_{j+1} | x_j\} | x_0\} \\ &= E\{w_j d_j' w_j | x_0\} = E\{w_j w_j' d_j | x_0\} = E\{w_j w_j' | x_0\} d_j \end{aligned} \tag{3.3}$$

and thus (3.2) becomes

$$\begin{aligned} d_{j+1} &= \Gamma_{j+1}[\Gamma_j^{-1} + E\{w_j w_j' | x_0\}/\Omega] d_j \\ &= \Gamma_{j+1}\Gamma_{j+1}^{-1} d_j = d_j, \end{aligned} \tag{3.4}$$

so that the update rule for the means may be dropped from the problem. The expected value term, $E\{w_j w_j' | x_0\}$ is evaluated using the modified, nonrandom means and variances of $d$.

The approximation described above can be used to convert the original problem into a sequence of stochastic open-loop problems, each of which involves minimization of the expected value of a quadratic objective function subject to a stochastic difference equation in $x$ and to a deterministic variance update equation. The interaction between estimation and control is still present, however, in the modified problem since policy variables affect future values of both $\Gamma$ and $x$.

## 4. Results

Using the approximation described in Section 3, the mathematical statement of the open-loop problem beginning in the first period is as follows. Choose strategy rules for policy variables $u_0, u_1, \ldots u_{N-1}$ so as to minimize

$$J = E\left\{\sum_{j=1}^{N} \tfrac{1}{2} q_j x_j^2 + \tfrac{1}{2} r_j u_{j-1}^2 + s_j x_j + t_j u_{j-1} | x_0\right\} \tag{4.1}$$

subject to

$$x_{j+1} = \mathbf{a} x_j + \mathbf{b} u_j + \mathbf{c} + \varepsilon_j = \mathbf{d} w_j + e_j, \qquad j = 0, \ldots, N-1 \tag{4.2}$$

and

$$\Gamma_{j-1}^{-1} = \Gamma_j^{-1} + E\{w_j w_j' | x_0\}/\Omega, \tag{4.3}$$

where $\mathbf{a}, \mathbf{b}, \mathbf{c}$, (and $\mathbf{d}$) are random variables with nonrandom means $a, b, c$ (and $d$) and conditional covariance matrix $\Gamma_j$, and $x_0$ and $\Gamma_0$ are given.

Since the variance constraint is deterministic, it will be handled by introducing matrices of Lagrangean multipliers, $M_j$, and forming the augmented objective function.

$$V = J + \frac{1}{2}\sum_{j=1}^{N} \operatorname{tr}[M_j(\Gamma_j^{-1} - \Gamma_j^{-1} - E\{w_{j-1}w_{j-1}'|x_0\}/\Omega], \tag{4.4}$$

which may be rewritten as

$$V = \sum_{j=1}^{N} (E\{W_j|x_0\} + \tfrac{1}{2}\operatorname{tr}[(M_{j-1} - M_j)\Gamma_{j-1}^{-1}]) + \tfrac{1}{2}\operatorname{tr}[M_N\Gamma_N^{-1} - M_0\Gamma_0^{-1}], \tag{4.5}$$

where

$$W = \tfrac{1}{2}q_jx_j^2 + \tfrac{1}{2}r_ju_{j-1}^2 + s_jw_j + t_ju_{j-1} - \tfrac{1}{2}\operatorname{tr}[M_jw_{j-1}w_{j-1}']/\Omega. \tag{4.6}$$

The constraint (4.2) is stochastic and cannot be handled through Lagrangean multipliers. Instead, it will be incorporated into the solution through the use of dynamic programming. The algorithm to minimize $J$ subject to (4.2) and (4.3) is represented by

$$V_N = \tfrac{1}{2}\operatorname{tr}[M_N\Gamma_N^{-1} - M_0\Gamma_0^{-1}] = V_N^*, \tag{4.7}$$

$$V_j = E\{V_{j+1}^* + W_{j+1}|x_j\} + \tfrac{1}{2}\operatorname{tr}[(M_j - M_{j+1})\Gamma_j^{-1}] \qquad j = N-1, \dots, 0. \tag{4.8}$$

$$V_j^* = \text{minimum of } V_j \text{ with respect to } u_j \text{, subject to (4.2)} \qquad j = N, \dots, 0, \tag{4.9}$$

$$\partial V/\partial \Gamma_j^{-1} = 0 \qquad j = N, \dots, 1, \tag{4.10}$$

and

$$\partial V/\partial M_j = 0 \qquad j = N, \dots, 0. \tag{4.11}$$

The form of the expressions $V_j$ is the same for all $j$. That is,

$$\begin{aligned} V_j = {} & E\{\tfrac{1}{2}k_{j+1}x_{j+1}^2 + \tfrac{1}{2}r_{j+1}u_j^2 + g_{j+1}x_{j+1} + t_{j+1}u_j \\ & - \tfrac{1}{2}\operatorname{tr}[M_{j+1}w_jw_j']/\Omega|x_0\} \\ & + \tfrac{1}{2}\operatorname{tr}[(M_j - M_{j+1})\Gamma_j^{-1}] \\ & + (\text{terms not involving } x, u \text{ or } \Gamma_j), \end{aligned} \tag{4.12}$$

for appropriate choice of $k_{j+1}$ and $g_{j+1}$. To see this note that the form of (4.12) is certainly true for $j = N - 1$, with $k_N = q_N$ and $g_N = s_N$. By induction, (4.12) can be shown to be correct for all $j$. For $V_j$ can be rewritten as

$$\begin{aligned} V_j = {} & \tfrac{1}{2}[k_{j+1}(a^2 + \Gamma_j^{aa}) - M_{j+1}^{aa}/\Omega]x_j^2 \\ & + \tfrac{1}{2}[k_{j+1}(b^2 + \Gamma_j^{bb}) - M_{j+1}^{bb}/\Omega + r_{j+1}]u_j^2 \\ & + \tfrac{1}{2}[k_{j+1}(c^2 + \Gamma_j^{cc}) - M_{j+1}^{cc}/\Omega] \\ & + [k_{j+1}(ab + \Gamma_j^{ab}) - M_{j+1}^{ab}/\Omega]x_ju_j \\ & + [k_{j+1}(bc + \Gamma_j^{bc}) - M_{j+1}^{bc}/\Omega]u_j \\ & + [k_{j+1}(ac + \Gamma_j^{ac}) - M_{j+1}^{ac}/\Omega]x_j \\ & + ag_{j+1}x_j + [bg_{j+1} + t_{j+1}]u_j + cg_{j+1} \\ & + \tfrac{1}{2}\operatorname{tr}[(M_j - M_{j+1})\Gamma_j^{-1}] \\ & + (\text{terms not involving } x, u \text{ or } \Gamma_j). \end{aligned} \tag{4.13}$$

The superscript letters on $\Gamma$ and $M$ specify particular elements of those matrices, e.g., $\Gamma^{bc}$ is the covariance between $b$ and $c$.

Differentiating (4.13) with respect to $u_j$ and setting the result equal to zero yields.

(4.14)
$$\begin{aligned}\partial V_j/\partial u_j &= [k_{j+1}(b^2 + \Gamma_j^{bb}) - M_{j+1}^{bb}/\Omega + r_{j+1}]u_j \\ &\quad + [k_{j+1}(ab + \Gamma_j^{ab}) - M_{j+1}^{ab}/\Omega]x_j \\ &\quad + [k_{j+1}(bc + \Gamma_j^{bc}) - M_{j+1}^{bc}/\Omega] \\ &\quad + bg_{j+1} + t_{j+1}. \\ &= 0,\end{aligned}$$

whence the optimal $u_j$ is given by the feedback rule

(4.15)
$$u_j^* = -(F_j x_j + f_j)/H_j,$$

where

(4.16)
$$H_j = k_{j+1}(b^2 + \Gamma_j^{bb}) - M_{j+1}^{bb}/\Omega + r_{j+1},$$

(4.17)
$$F_j = k_{j+1}(ab + \Gamma_j^{ab}) - M_{j+1}^{ab}/\Omega,$$

and

(4.18)
$$f_j = k_{j+1}(bc + \Gamma_j^{bc}) - M_{j+1}^{bc}/\Omega + bg_{j+1} + t_{j+1}.$$

Substituting the optimal $u_j^*$ back into $V_j$ gives $V_j^*$:

(4.19)
$$\begin{aligned}V_j^* &= \tfrac{1}{2}[k_{j+1}(a^2 + \Gamma_{j+1}^{aa}) - M_{j+1}^{aa}/\Omega - F_j^2/H_j]x_j^2 \\ &\quad + [ag_{j+1} + k_{j+1}(ac + \Gamma_j^{ac}) - M_{j+1}^{ac}/\Omega - F_j f_j/H_j]x_j \\ &\quad + (\text{terms not involving } x \text{ or } u).\end{aligned}$$

The derivative, $\partial V/\partial\Gamma_j^{-1}$, which is called for by (4.10), is identical with the derivative, $\partial E\{V_j|x_0\}/\partial\Gamma_j^{-1}$, since $V_j$ is the only portion of $V$ involving $\Gamma_j$. The expected value of $V_j$ is

(4.20)
$$\begin{aligned}E\{V_j|x_0\} &= \tfrac{1}{2}\operatorname{tr}[k_{j+1}E\{w_j w_j'|x_0\}\Gamma_j + (M_j - M_{j+1})\Gamma_j^{1}] \\ &\quad + (\text{terms not involving } \Gamma_j).\end{aligned}$$

Hence, setting the derivative of $V$ with respect to $\Gamma_j^{-1}$ equal to zero[4] and noting that for any scalar-valued function $f$,

(4.21)
$$\partial f/\partial\Gamma_j^{-1} = -\Gamma_j(\partial f/\partial\Gamma_j)\Gamma_j,$$

yields the following difference equation in $M$:

(4.22)
$$M_j = M_{j+1} + \Gamma_j(k_{j+1}E\{w_j w_j'|x_0\})\Gamma_j,$$

with $M_N = 0$.

[4] See [5] for a more detailed discussion of matrix derivatives.

Finally, the expression for $V_j^*$ given by (4.19) is used to derive $V_{j-1}$:

$$(4.23)\quad V_{j-1} = E\{V_j^* + W_j | x_{j-1}\} + \tfrac{1}{2}\operatorname{tr}[M_{j-1} - M_j)\Gamma_{j-1}^{-1}]$$
$$= E\{\tfrac{1}{2}k_j x_j^2 + \tfrac{1}{2}r_j u_{j-1}^2 + g_j x_j + t_j u_{j-1} - \tfrac{1}{2}\operatorname{tr}[M_j w_{j-1} w'_{j-1}]/\Omega | x_0\}$$
$$+ \tfrac{1}{2}\operatorname{tr}[(M_{j-1} - M_j)\Gamma_{j-1}^{-1}$$
$$+ \text{(terms not involving } x, u \text{ or } \Gamma_{j-1}),$$

where

$$(4.24)\qquad k_j = q_j + k_{j+1}(a^2 + \Gamma_j^{aa}) - M_{j+1}^{aa}/\Omega - F_j^2/H_j,$$

$$(4.25)\qquad g_j = s_j + ag_{j+1} + k_{j+1}(ac + \Gamma_j^{ac}) - M_{j+1}^{ac}/\Omega - F_j f_j/H_j,$$

The system of equations, (4.15)–(4.18), (4.22), (4.24) and (4.25), along with the two constraints, (4.2) and (4.3), characterize an optimal solution for the open-loop problem beginning in period 1. The quantities $k_j$ and $g_j$, together with $r_j$ and $t_j$, comprise the coefficients of an intertemporal objective function, in which the evaluation of various values for the current policy variable reflects not only the current impact but also the future impact through the difference equation in $x$, (4.2). Like $k_j$ and $g_j$, the matrix $M_j$ also reflects the future, but through the dynamic variance update equation (4.3). If equation (4.2) is not dynamic, that is, if the parameter $a$ is known to be equal to zero, then $k_j = q_j$ and $g_j = s_j$, so that the intertemporal objective function reflects only the current impact of current policies. Similarly, if the variance update equation ceases to be dynamic because the prior variance matrix is identically zero, then $M_j$ will also be zero and current policies will have no future impact through learning.

The matrix $M_j$ is, by the theory of Lagrangean multipliers, the imputed price of $\Gamma_j^{-1}$, which may be thought of as the stock of information available at period $j$. Since $M_j$ is equal to the sum of positive semidefinite terms, it will also be positive semi-definite, and will be smaller (i.e., less positive definite) for larger stocks of information of equivalently for larger values of $\Gamma_j$, representing diminishing marginal returns to increased information. The price of information is also larger for larger values of $k_{j+1}$, the cost in the intertemporal objective function of imperfect knowledge about $a$, $b$ and $c$, and consequently about $x_{j+1}$.

Except for equation (4.21), matrix $M$ never appears alone in the equations characterizing an optimal solution, but is always multiplied by $1/\Omega$, a quantity which may be interpreted as the amount of information ultimately available from the basic dynamic equation, (4.2). A large $\Omega$, and consequently small value for $1/\Omega$, indicates that equation (4.2) is so noisy that additional observations on $x$ and $u$ can provide very little improvement in the estimates of $a$, $b$ and $c$. The expression $M/\Omega$, therefore, may be looked upon as the value of estimating. This value-of-estimating term appears with a negative coefficient in $V_j$, the expression which is minimized by appropriate choice of $u_j$. A large value of $M/\Omega$ will offset costs incurred by choosing a policy which is non-optimal from a purely control point of view.

Models in which the parameters are treated as known quantities and models where uncertainty is present but learning is ignored have solutions which are special cases of the adaptive learning solution presented above. In a certainty equivalence model, where the parameters $a$, $b$ and $c$ are considered fixed numbers, the optimal sequence of strategy rules are identical to those given above, except that $\Gamma$ and $M$ are identically zero. For the model where the uncertain parameters are modelled as independently and identically distributed in each period, $\Gamma$ in every period equals the prior variance matrix and $M$ is zero.

A comparison of the optimal first period policy rules of the three different types is quite straightforward from a short-run or static point of view. That is, assuming $k_1$ and $g_1$ to be the same for all three cases, the difference between the uncertainty aversion rule (unknown parameters but no learning) and the certainty equivalence rule lies merely in the addition of a variance term to $H_0$, $F_0$ and $f_0$, the quantities which comprise the strategy rule for $u_0$. The rule for the adaptive learning case includes a further term subtracted from $H_0$, $F_0$, and $f_0$, which counteracts, to some extent, the effect of uncertainty. At first glance then, and for simplicity assuming that only the parameter $b$ is uncertain, it appears that the uncertainty aversion policy rule is more conservative (since $1/H_0$ is smaller) than the certainty equivalence rule, and the adaptive learning rule is more aggressive and does more learning than the latter.

A short-run evaluation, however, of the differences among the three cases is misleading. If the model is dynamic then, as pointed out earlier, $k_1$ and $g_1$ will reflect future behavior of the model, and in general this will differ for the three different cases. Thus it is possible, and indeed is borne out by the numerical example which follows, that the first period adaptive learning policy will actually be *more* conservative than the uncertainty aversion policy. That is, since $k_1$ and $g_1$ in the adaptive learning case reflect the fact that more will be known later even with no active experimentation, the optimal first period policy may be to do almost nothing, then make up for it later when the effect of control action is better known.

## 5. Example

To illustrate the behavior of the adaptive learning model optimal first period policies are calculated for a simple dynamic model. To simplify matters, only parameter $b$ is considered to be unknown, with prior mean of $-0.5$ and varying prior variances. Parameters $a$ and $c$ are known to be 0.7 and 3.5 respectively,[5] and the noise variance, $\Omega$, is 0.2. The objective function is written as a sum of squared deviations from goals on $x$ and $u$,

$$J = E\left\{\sum_{j=1}^{N} \tfrac{1}{2}q(x_j - \hat{x})^2 + \tfrac{1}{2}r(u_{j-1} - \hat{u})^2 | x_0\right\} \tag{5.1}$$

where the values of $q$, $r$, $x$ and $u$ are the same for all periods. Note that this form of the objective function can be rewritten to have the same form as (2.9), except for a constant term.

The certainty equivalence, uncertainty aversion and adaptive learning policies for the first period of a four period problem are shown in Table 1 for various goals, prior variances and $q:r$ ratios. As can be seen, there are clearly some combinations of parameters, variances, and objective functions which cause the adaptive policy to be less aggressive and do less learning than the uncertainty aversion policy.

[5] See [4] for an application to the problem of optimally controlling inflation and unemployment where $a$, $b$, and $c$ are all unknown.

These instances are marked with asterisks in the tables. From casual inspection it is clear that some sort of pattern prevails in the location of these paradoxical results. Since none of them appear when the prior variance equals 2.0, it seems that if the initial uncertainty is relatively great, some active learning must be carried out immediately. It also seems from Table 1 that the strength of the paradoxical behavior of the adaptive policy is related to the relative weights on $x$ and $u$. That is, higher $q:r$ ratios lead to relatively less aggressive initial policies. One possible reason for this is that the higher the level of $q$ relative to $r$, the greater will be the *intertemporal* ratio $k:r$ in early periods as compared with later. This means that the unknown effect of policy action on $x$ is relatively more costly at the beginning of the planning period than at the end when both the $k:r$ ratio and the variance will be lower, so the initial policy tends closer to a do-nothing policy.

The longer is the planning horizon the more learning can occur. This suggests that the optimal first period policy for long planning horizon would be relatively more conservative than for short horizons. The results shown in Table 2, which shows first period policies for the uncertainty aversion and adaptive learning cases with varying lengths of the planning period, seem to bear this out.

TABLE 1

FIRST PERIOD POLICIES FOR DIFFERENT GOALS AND PRIOR $b$ VARIANCES

Horizon = 4

$a = 0.7$ $b = -0.5$ $c = 3.5$

$\Omega = 0.2$ $\Gamma^{aa} = \Gamma^{cc} = 0.0$

| | | $\Gamma_0^{bb} = 0.5$ | | $\Gamma_0^{bb} = 1.0$ | | $\Gamma_0^{bb} = 2.0$ | |
|---|---|---|---|---|---|---|---|
| $q:r$ | Cert. Equiv. | Uncert. Averse | Adaptive | Uncert. Averse | Adaptive | Uncert. Averse | Adaptive |
| | | | Goals: $\hat{x} = 0, \hat{u} = 0$ | | | | |
| 0:5 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 1:5 | 1.201 | 1.046 | 1.082 | 0.925 | 0.973 | 0.751 | 0.820 |
| 5:5 | 3.562 | 2.524 | 2.449* | 1.929 | 1.923* | 1.302 | 1.446 |
| 5:1 | 5.821 | 3.578 | 3.056* | 2.489 | 2.316* | 1.530 | 1.759 |
| 5:0 | 7.000 | 4.017 | 3.146* | 2.688 | 2.427* | 1.601 | 1.880 |
| | | | Goals: $\hat{x} = 1, \hat{u} = 0$ | | | | |
| 0:5 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 1:5 | 0.985 | 0.858 | 0.898 | 0.760 | 0.815 | 0.617 | 0.695 |
| 5:5 | 2.869 | 2.051 | 2.033* | 1.574 | 1.626 | 1.066 | 1.249 |
| 5:1 | 4.472 | 2.874 | 2.528* | 2.020 | 1.973* | 1.251 | 1.529 |
| 5:0 | 5.000 | 3.206 | 2.596* | 2.178 | 2.060* | 1.308 | 1.618 |
| | | | Goals: $\hat{x} = 0, \hat{u} = 1$ | | | | |
| 0:5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1:5 | 2.029 | 1.767 | 1.788 | 1.564 | 1.586 | 1.269 | 1.307 |
| 5:5 | 4.053 | 2.871 | 2.751* | 2.194 | 2.141* | 1.480 | 1.592 |
| 5:1 | 5.989 | 3.676 | 3.124* | 2.557 | 2.361* | 1.572 | 1.789 |
| 5:0 | 7.000 | 4.017 | 3.146* | 2.688 | 2.427* | 1.601 | 1.880 |
| | | | Goals: $\hat{x} = 1, \hat{u} = 1$ | | | | |
| 0:5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1:5 | 1.814 | 1.580 | 1.606 | 1.398 | 1.429 | 1.135 | 1.182 |
| 5:5 | 3.360 | 2.397 | 2.332* | 1.839 | 1.842 | 1.245 | 1.397 |
| 5:1 | 4.640 | 2.972 | 2.595* | 2.088 | 2.018* | 1.293 | 1.560 |
| 5:0 | 5.000 | 3.206 | 2.596* | 2.178 | 2.060* | 1.308 | 1.618 |

* indicates examples where the adaptive policy initially learns less than the uncertainty aversion policy

TABLE 2

FIRST PERIOD POLICIES FOR DIFFERENT HORIZON LENGTHS

$N$ = Horizon
Goals = 0
$a = 0.7$  $b = -0.5$  $c = 3.5$
$\Omega = 0.2$  $\Gamma_0^{bb} = 0.5$  $x_0 = 0.0$

| $q:r$ | $N=2$ Uncert. | Adapt. | $N=4$ Uncert. | Adapt. | $N=8$ Uncert. | Adapt. | $N=16$ Uncert. | Adapt. |
|---|---|---|---|---|---|---|---|---|
| 0:5 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 1:5 | 0.613 | 0.622 | 1.046 | 1.082 | 1.362 | 1.394 | 1.434 | 1.460 |
| 5:5 | 1.712 | 1.740 | 2.524 | 2.449* | 2.959 | 2.688* | 3.016 | 2.705* |
| 5:1 | 2.691 | 2.682* | 3.578 | 3.056* | 3.957 | 3.083* | 3.987 | 3.084* |
| 5:0 | 3.154 | 3.138* | 4.017 | 3.146* | 4.354 | 3.147* | 4.375 | 3.147* |

Finally, Table 3 compares the results obtained through the structural approximation described in this paper to those obtained in Prescott's paper[6] through a numerical approximation. The model for which the policies are computed is static, with $a$ and $c$ both equal to zero, and the objective function has goals and weights only on $x$. From the table it appears that the adaptive approximation is reasonable when the ratio of prior standard deviation to prior means is less than about 1.5. This suggests that the paradoxical results shown in Tables 1 and 2 would also be valid for a true optimal solution, since the prior variances are relatively small compared with the prior mean of the unknown parameters.

## 6. CONCLUSIONS

The general linear decision model with unknown parameters in which learning is explicitly incorporated is in general insolvable because of the interaction between the control and estimation roles of the policy variables. The approximation

TABLE 3

COMPARISON OF STRUCTURAL AND NUMERICAL APPROXIMATIONS FOR STATIC PROBLEM

Horizon = 4
$q:r = 1:0$  $\Omega_{bb} = 1.0$
$a = c = 0.0$  $\Gamma_0^{bb} = 1.0$

| $b$ | Goals: $\hat{x}=1, \hat{u}=0$ Uncert. Averse | Adapt. Approx. | Numer. Approx. | Goals: $\hat{x}=4, \hat{u}=0$ Uncert. Averse | Adapt. Approx. | Numer. Approx. |
|---|---|---|---|---|---|---|
| 0.0 | 0.00 | 0.00 | 0.56 | 0.00 | 0.00 | 2.00 |
| 0.2 | 0.19 | 0.22 | 0.69 | 0.77 | 1.63 | 2.30 |
| 0.4 | 0.34 | 0.54 | 0.72 | 1.38 | 2.83 | 2.55 |
| 0.7 | 0.47 | 0.70 | 0.68 | 1.88 | 2.62 | 2.63 |
| 1.0 | 0.50 | 0.65 | 0.58 | 2.00 | 2.41 | 2.54 |
| 1.4 | 0.47 | 0.54 | 0.50 | 1.89 | 2.11 | 2.20 |
| 2.0 | 0.40 | 0.42 | 0.41 | 1.60 | 1.72 | 1.76 |
| 3.0 | 0.30 | 0.30 | 0.30 | 1.20 | 1.25 | 1.25 |
| 4.0 | 0.24 | 0.24 | 0.24 | 0.94 | 0.97 | 0.97 |
| 5.0 | 0.19 | 0.19 | 0.19 | 0.77 | 0.78 | 0.78 |

[6] See Table 2 in Edward Prescott's paper, "The Multiperiod Control Problem Under Uncertainty," *Econometrica*, forthcoming.

developed in this paper is designed to retain some of the stochastic nature of the original problem along with the interaction between estimation and control, but to permit an actual solution to be derived. The approach developed in this paper treats a general linear dynamic model where any or all of the parameters may be unknown. Although the discussion in this paper is confined to a model consisting of a single equation, the extension to multiequation models is perfectly straightforward with appropriate notation.

One advantage of the approach used is that the solution involves Lagrangean multipliers which can readily be interpreted in terms of such economic concepts as price of information and value of estimating. Furthermore, the mathematical form of the solution is a straightforward generalization of the solutions for the certainty equivalence problem where all parameters are considered known, and the uncertainty aversion problem where the parameters are unknown but learning is ignored. Lastly, although the comparisons given in this paper are rather limited, it appears that the approximation compares favorably with numerical results when the prior variances are not too large.

A common feeling with regard to adaptive models of this type is that if it is possible to learn, i.e., to experiment so as to improve future estimates, the optimal policy will indeed be more aggressive and learn more than a policy based on a model where the level of uncertainty about the parameters is the same for all periods. One of the results of this paper, however, is that the adaptive policy is often more conservative than the non-adaptive policy. That is, the future gains from actively experimenting may not offset the cost incurred by the uncertain current effect of such a choice of policy variables. The optimal policy may well be to do very little initially, thus insuring that whatever the unknown effect is it is quite small, and do all necessary control action later when more will be known.

*University of Maryland*

## REFERENCES


[1] Aoki, Masanao. *Optimization of Stochastic Systems*. New York: Academic Press, 1967.

[2] Brainard, William C., "Uncertainty and the Effectiveness of Policy," *American Economic Association Papers and Proceedings*, Vol. 57 (May 1967).

[3] Holt, C. C., F. Modigliani, and J. F. Muth, "Derivation of a Linear Decision Rule for Production and Employment," *Management Science*, Vol. 2 (June 1956).

[4] MacRae, C. Duncan and Elizabeth Chase MacRae, "Adaptive Control of Inflation and Unemployment," *NEREM Record*, (November 1970).

[5] MacRae, Elizabeth Chase, "Matrix Derivatives with an Application to the Analysis of Covariance Structures," Special Studies Paper #20, Federal Reserve Board, (September 1971).

[6] MacRae, Elizabeth Chase, "Optimal Estimation and Control: A Structural Approximation," Special Studies Paper #27, Federal Reserve Board, (August 1972).

[7] MacRae, Elizabeth Chase and C. Duncan Macrae, "Linear Decision with Uncertain Parameters," Special Studies Paper #28, Federal Reserve Board, (forthcoming).

[8] Murphy, W. J., "Optimal Stochastic Control of Discrete Linear Systems with Unknown Gain," *IEEE Trans. Automat. Contr.*, Vol. AC-13 (August 1968).

[9] Theil, Henri, *Optimal Decision Rules for Government and Industry*, Chicago: North-Holland, 1964.

[10] Tse, Edison and Michael Athans, "Adaptive Stochastic Control for a Class of Linear Systems," *IEEE Trans. Automat. Contr.*, Vol. AC-17 (February 1972).