A SOLUTION TO OPTIMAL CONTROL OF LINEAR SYSTEMS WITH UNKNOWN PARAMETERS

Gregory C. Chow

Econometric Research Program
Research Memorandum No. 157
December 1973

The research described in this paper was supported by NSF Grant GS 32003X.

Econometric Research Program
PRINCETON UNIVERSITY
207 Dickinson Hall
Princeton, New Jersey

A SOLUTION TO OPTIMAL CONTROL OF LINEAR SYSTEMS WITH UNKNOWN PARAMETERS

Gregory C. Chow *

1. Introduction

In the study of optimal economic policy using a linear econometric model and a quadratic welfare function, the parameters of the model are often assumed to be known for certain. Under this assumption, the solution in the form of an optimal feedback control equation can be easily obtained. Although one recognizes that in a realistic situation, the parameters of an econometric model are never known for certain, he might still apply the above solution, using a set of estimates of the parameters as if they were the true values, if he believes that it is a good approximation to optimal policy. Such a procedure is well known to be a certainty equivalence solution.

In a recent paper, Chow (1973b), I have presented a method of obtaining the optimal feedback control equations and the associated welfare costs under the assumptions that the uncertainty in the parameters is expressed in the posterior density computed by Bayesian methods using data available up to the time of decision and that this posterior density shall not be revised in the future for the derivation of the current policy. Because future learning about the model is not explicitly taken into account in the design of the current policy, the above method is not truly optimal. However, if the sample period is long as compared with the planning period, this method will probably be close to being optimal. Furthermore, when this solution is compared with the certainty equivalence solution in order to study the effects of uncertainty on the optimal policy and the associated expected welfare cost, it will provide upper limits to the impact of uncertainty. Presumably, the presence of uncertainty will decrease the value of the optimal policy, and the possibility for learning will make the truly optimal policy superior to the above policy, thus increasing the value of control.

The purpose of this paper is to present a solution to optimal control when the possibility of learning is taken into account in the design of the current policy. In the control literature, this problem is often referred to as the dual control problem because control is to serve the dual purpose of improving the system performance and of learning more about the system for the sake of future control. Numerous approximate solutions have been suggested.[1] The solution of this paper appears to be the simplest in conception, and yet it incorporates all theoretical elements in the calculations. It accepts no compromise on any theoretical relationships, such as the assumption that the expectation of a function is the function of the expectation, or that the expectation of a minimum is the minimum of expectation, or that a random variable equals its mean value, or that a variable in period $t+1$ equals its value in period t, or that learning can take place for no more than two periods ahead. It contains a logical structure which brings out clearly the effect of learning on the optimization process and enables the effect to be measured numerically. It provides useful contrasts to the certainty equivalence solution and the solution for unknown parameters without learning, being a natural generalization of these two solutions.

We will set up the problem and describe the method of solution in section 2. This method will be compared in section 3 with the two other methods just mentioned, both in conceptual terms and in terms of computations. Two simpler, modified versions of the method will also be briefly described. They are simpler to compute, but they still take learning partially into account. Some numerical results using a simple one-equation model will be presented in section 4, to bring out the effects of learning on the optimal control solution. This paper is confined mainly to presenting the method and providing some illustrative calculations. A comprehensive study of the effect of learning on optimal policies using the method of this paper remains to be undertaken.

From the viewpoint of economics in general, other than the study of quantitative economic policy using econometric models, the content of this paper may also be relevant. Maximization is in the heart of economics. Most of economic theory assumes maximization to take place in a static situation under certainty. When generalizations are made in a multiperiod, dynamic situation under uncertainty, the present paper shows the nature and complexity of the solution. Its results are perhaps useful to the economist who wishes to reflect on the applicability of maximizing behavior to dynamic and stochastic economics, and to investigate into the form that a dynamic theory should take and the way it can be applied to the study of actual behavior.

2. Description of the Method

Assume a linear econometric model of the form

$$y_t = A\, y_{t-1} + C\, x_t + u_t \; . \tag{2.1}$$

Here y_t is a vector consisting of current dependent variables, lagged dependent variables (to convert an originally higher-order system into a first order system), current control variables (which are simply equal to the right-hand-side control variables x_t, with the corresponding row of A equal to 0), and possibly lagged control variables (which are explained by identities).[2] A and C are matrices with unknown elements, except for the rows corresponding to identities in the system. u_t is a random vector (having zero elements corresponding to the identities) with zero mean and unknown covariance matrix Σ, and is independent of u_s for $t \neq s$. There may be exogenous variables z_t in the system, with unknown coefficients B, but we have omitted Bz_t on the right-hand side of (2.1) to simplify the algebra in the following.[3]

The welfare cost for the planning period from 1 to T is assumed to be

$$(2.2)\quad W = \frac{1}{2}\sum_{t=1}^{T} y_t' K_{t,t} y_t + \sum_{t=1}^{T} \sum_{s<t} y_t' K_{t,s} y_s + \sum_{t=1}^{T} y_t' k_t + d$$

where $K_{t,s} = K'_{s,t}$, k_t and d are known constants. The function (2.2) is more general than what can be found in the control literature which usually assumes that $K_{t,s} = 0$ for $t \neq s$, or that welfare cost takes the form of a sum $\sum_t W_t(y_t)$. The standard treatment using dynamic programming makes this special assumption. We will treat the more general welfare cost function in this paper.

The problem is to choose $x_1, \ldots, x_T$ sequentially to minimize the expected welfare cost as of the beginning of period 1. Since the policy variables for later periods need not be chosen until the outcomes of earlier decisions are available, the problem should be solved by first minimizing out the control variables for the later periods, given the outcomes of earlier decisions, and proceeding successively backward in time until the control variable x_1 for the first period is to be chosen. The logical structure of the solution is thus. Given a welfare cost function $W = W(y_1, \ldots, y_T)$, we first eliminate x_T by minimizing the conditional expectation of W given all the data up to $T-1$, then eliminate x_{T-1} by minimizing the conditional expectation of the above minimum given all the data up to $T-2$, and so forth until we minimize, with respect to x_1 , a conditional expectation given the data at the end of time 0 :

$$(2.3) \qquad \min_{x_1} (E_0 \cdots (\min_{x_{T-2}} E_{T-3}(\min_{x_{T-1}} E_{T-2} (\min_{x_T} E_{T-1} W)))\cdots) .$$

To carry out the first minimization with respect to x_T we write

$$(2.4) \qquad E_{T-1}W = E_{T-1}[\tfrac{1}{2} y_T' H_{T,T}^T y_T + y_T'(\sum_{s=1}^{T-1} H_{T,s}^T y_s + h_T^T)]$$

$$+ \tfrac{1}{2} \sum_{t=1}^{T-1} y_t' H_{t,t}^T y_t + \sum_{t=1}^{T-1} \sum_{s<t} y_t' H_{t,s}^T y_s + \sum_{t=1}^{T-1} y_t' h_t^T + d_T$$

$$= E_{T-1} W_T + W_{NT}$$

where the function W has been decomposed into two parts, W_T and W_{NT} . All the terms involving y_T , which can be influenced by x_T , are included in W_T . The terms in W_{NT} are not affected by x_T . We have let $K_{t,s} = H_{t,s}^T$, $k_t = h_t^T$ and $d = d_T$ to facilitate generalization in a future step (following equation (2.9) below). To minimize $E_{T-1} W_T$, we substitute $Ay_{T-1} + Cx_T + u_T$ for y_T in W_T and take expectation:

$$(2.5) \quad E_{T-1} W_T = E_{T-1}[\tfrac{1}{2}(Ay_{T-1} + Cx_T + u_T)' H^T_{T,T}(Ay_{T-1} + Cx_T + u_T)$$

$$+ (Ay_{T-1} + Cx_T + u_T)' (\sum_{s=1}^{T-1} H^T_{T,s} y_s + h^T_T)]$$

$$= \tfrac{1}{2} y'_{T-1}(E_{T-1}A'H^T_{T,T}A + 2 E_{T-1} A'H^T_{T,T-1})y_{T-1} + \tfrac{1}{2}x'_T(E_{T-1}C'H^T_{T,T}C)x_T$$

$$+ x'_T[E_{T-1}(C'H^T_{T,T} A + C'H^T_{T,T-1})y_{T-1} + (E_{T-1}C')(\sum_{s=1}^{T-2} H^T_{T,s} y_s + h^T_T)]$$

$$+ y'_{T-1}(E_{T-1}A')(\sum_{s=1}^{T-2} H^T_{T,s} y_s + h^T_T) + E_{T-1} u'_T H^T_{T,T} u_T .$$

When taking expectations, we have adopted the Baysian view that the matrices A and C have a joint posterior density function at the end of $T-1$, and have assumed that u_T is distributed independently of them.

Minimization of (2.5) by differentiation with respect to x_T yields

$$(2.6) \quad \hat{x}_T = - (E_{T-1}C'H^T_{T,T}C)^{-1}[E_{T-1}(C'H^H_{T,T}A + C'H^T_{T,T-1})y_{T-1}$$
$$+ (E_{T-1}C')(\sum_{s=1}^{T-2} H^T_{T,s}y_s + h^T_T)]$$

(2.6) is a feedback control equation, determining the optimal policy for period T in terms of observations $y_1, y_2, \ldots, y_{T-1}$. These observations affect the posterior dentisy of A and C and thus the expectations involving them. Substituting the solution

(2.6) for x_T in (2.5) gives

(2.7) $$\underset{x_T}{\text{Min}}\ E_{T-1}W_T = \frac{1}{2} y'_{T-1}(E_{T-1}A'H^T_{T,T}A + 2\ E_{T-1}A'H^T_{T,T-1})y_{T-1}$$

$$- \frac{1}{2}[y'_{T-1}\ E_{T-1}(A'H^T_{T,T}C + H^{T'}_{T,T-1}C) + (\sum_{s=1}^{T-2} y'_s H^{T'}_{T,s} + h^{T'}_T)(E_{T-1}C)][E_{T-1}C'H^T_{T,T}C]^{-1} \times$$

$$[E_{T-1}(C'H^T_{T,T}A + C'H^T_{T,T-1})y_{T-1} + (E_{T-1}C')(\sum_{s=1}^{T-2} H^T_{T,s}\,y_s + h^T_T)]$$

$$+ y'_{T-1}(E_{T-1}A')(\sum_{s=1}^{T-2} H^T_{T,s}\,y_s + h^T_T) + \frac{1}{2}\ E_{T-1}\ u'_T H^T_{T,T}\,u_T$$

$$= \frac{1}{2} y'_{T-1}[E_{T-1}A'H^T_{T,T}A + 2E_{T-1}A'H^T_{T,T-1} - E_{T-1}(A'H^T_{T,T}C + H^{T'}_{T,T-1}C)(E_{T-1}C'H^T_{T,T}C)^{-1} \times$$
$$E_{T-1}(C'H^T_{T,T}A + C'H^T_{T,T-1})]\,y_{T-1}$$

$$+ y'_{T-1}[(E_{T-1}A') - E_{T-1}(A'H^T_{T,T}C + H^{T'}_{T,T-1}C)(E_{T-1}C'H^T_{T,T}C)^{-1}(E_{T-1}C')](\sum_{s=1}^{T-2} H^T_{T,s}\,y_s + h^T_T)$$

$$- \frac{1}{2}(\sum_{s=1}^{T-2} y'_s\ H^{T'}_{T,s} + h^{T'}_T)(E_{T-1}C)(E_{T-1}C'H^T_{T,T}C)^{-1}(E_{T-1}C')(\sum_{s=1}^{T-2} H^T_{T,s}y_s + h^T_T) + \frac{1}{2}E_{T-1}u'_T H^T_{T,T}u_T$$

The essense of the method of this paper is to approximate (2.7) by a quadratic function in $y_{T-1}, y_{T-2}, \ldots, y_1$. This quadratic function can then be combined with W_{NT} in (2.4) to yield $\underset{x_T}{\text{Min}}\ E_{T-1}W$, which will be quadratic in $y_{T-1}, y_{T-2}, \ldots, y_1$. One then proceeds to minimize $E_{T-2}(\underset{x_T}{\text{Min}}\, E_{T-1}W)$ with respect to x_{T-2}, following the steps from (2.4) on, with T-1 replacing T in all the derivations. To be specific, let

$$\underset{x_T}{\text{Min}}\, E_{T-1}W_T = \frac{1}{2}\sum_{t=1}^{T-1} y_t' Q_{t,t}^T y_t + \sum_{t=1}^{T-1}\sum_{s<t} y_t' Q_{t,s}^T y_s + \sum_{t=1}^{T-1} y_t' q_t^T + r_T . \tag{2.8}$$

Discussion of the choice of this quadratic approximation will be postponed until the next paragraph. Combining the right-hand side of (2.8) with W_{NT} in (2.4), one gets

$$\underset{x_T}{\text{Min}}\, E_{T-1}W = \frac{1}{2}\sum_{t=1}^{T-1} y_t' H_{t,t}^{T-1} y_t + \sum_{t=1}^{T-1}\sum_{s<t} y_t' H_{t,s}^{T-1} y_s + \sum_{t=1}^{T-1} y_t' h_t^{T-1} + d_{T-1} \tag{2.9}$$

where

$$H_{i,j}^{T-1} = H_{i,j}^T + Q_{i,j}^T \qquad (i=1,\ldots,T-1;\ j\le i)$$

$$h_i^{T-1} = h_i^T + q_i^T \tag{2.10}$$

$$d_{T-1} = d_T + r_T .$$

Once (2.9) is obtained, it can be treated in the same way as (2.4) with T replacing T-1. Thus (2.9) will be decomposed into two parts, W_{T-1} and $W_{N(T-1)}$, the former involving y_{T-1} while the latter does not. $E_{T-2}W_{T-1}$ will then be minimized with respect to x_{T-1}, yielding results analogous to (2.5) and (2.7). The analog of (2.7), namely $\underset{x_{T-1}}{\text{Min}}\ E_{T-2}\ W_{T-1}$, will be approximated by a quadratic function with coefficients $Q_{t,s}^{T-1}$, q_t^{T-1} and r_{T-1}. This quadratic function will be combined with $W_{N(T-1)}$ to yield $\underset{x_{T-1}}{\text{Min}}\ E_{T-2}(\underset{x_t}{\text{Min}}\ E_{T-1}W)$ as in (2.9). The coefficients of the last quadratic function are obtained by recursion formulas (2.10) with $T-1$ replacing T. Now one is back to minimizing the expectation of a quadratic function in the form of (2.4). The process will continue until one minimizes, with respect to x_1, the conditional expectation of a quadratic function in y_1, namely, $\frac{1}{2}\, y_1'\, H_{1,1}^1\, y_1 + y_1' h_1^1$.

To return to the problem of finding a quadratic approximation of (2.7), it is necessary to choose a tentative path $(y_1^o, \ldots, y_{T-1}^o)$ around which a second-order Taylor expansion of (2.7) will be numerically computed. Any good approximation to the solution of the optimal control problem of this paper can be used for this purpose. In particular, the solution of a previous paper (Chow, 1973b) can be used. In that paper, I have employed the method of solution as described above, <u>except</u> that all conditional expectations E_t are treated as E_o. Thus the possibility of learning (revising the posterior density of A and C) is ignored in deriving the optimal policy for x_1. When this approximation is taken, and under the assumption that $K_{t,s}$ in the welfare function (2.2) is zero for $t \neq s$, (2.6) and (2.7) will be reduced respectively to (for $1 \leq t \leq T$)

$$(2.11) \quad \hat{x}_t = -(E_o C'H_t C)^{-1} [(E_o C'H_t A) y_{t-1} + (E_o C')h_t]$$

(where $H_t \equiv H^t_{t,t}$ and $h_t \equiv h^t_t$) and

$$(2.12) \quad \underset{x_t}{\text{Min}}\ E_{t-1}W_t = \frac{1}{2} y'_{t-1}[E_o A'H_t A - (E_o A'H_t C)(E_o C'H_t C)^{-1}(E_o C'H_t A)]y_{t-1}$$
$$+ y'_{t-1}[E_o A' - (E_o A'H_t C)(E_o C'H_t C)^{-1} (E_o C')] h_t$$
$$- \frac{1}{2} h'_t(E_o C)(E_o C'H_t C)^{-1} (E_o C')h_t + \frac{1}{2} E_o u'_t H_t u_t \ .$$

(2.12) is truly a quadratic function of y_{t-1}, and need not be approximated. Thus, the coefficients of the quadratic function (2.8) will be reduced to

$$(2.13) \quad Q_{t-1} \equiv Q^t_{t-1,t-1} = E_o A'H_t A - (E_o A'H_t C)(E_o C'H_t C)^{-1}(E_o C'H_t A)$$

$$q_{t-1} \equiv q^t_{t-1} = [E_o A' - (E_o A'H_t C)(E_o C'H_t C)^{-1}(E_o C')] h_t$$

$$r_t = -\frac{1}{2} h'_t(E_o C)(E_o C'H_t C)^{-1} (E_o C')h_t + \frac{1}{2} E_o u'_t H_t u_t$$

$$Q^t_{i,j} = 0 \ ; \quad q^t_i = 0 \qquad (i < t-1,\ j \neq i)$$

The joint density of A and C, based on data up to any time t, can be obtained by Bayesian method. The first two moments of this density, which can be used to compute the expectations required in (2.13), are explicitly given in Chow (1973b).[4] One then uses (2.13) together with (2.10) to compute H_t, Q_t, h_t and d_t backward in time, $t = T-1, T-2, \ldots, 1$. Finally, $\hat{x}_1$ is obtained by a formula analogous to (2.11). This $\hat{x}_1$ will be employed to generate y_1, using the model (2.1), with E_oA and E_oC replacing A and C respectively and preferably but not necessarily including the stochastic disturbance u_1. $\hat{x}_2$ will be obtained from y_1 and the corresponding optimal control equation, and so forth. Thus a tentative path $(y_1^o, \ldots, y_{T-1}^o)$ can be obtained for the quadratic approximation of (2.7). Once the method described in this section is applied to obtain a better $\hat{x}_1$, one can use it to generate a second tentative path, possibly improving upon the previous feedback control equations in the calculations by stochastic simulations of an open-loop policy. Better $\hat{x}_1$ and better feedback control equations could always be used again to find better tentative paths, and vice-versa.

3. Comparison with Other Methods

In this section, the method of section 2 will be compared with two simpler, and less nearly optimal, methods in terms of the structure of the solution. The comparison will provide interesting interpretations of the various calculations required in obtaining the solution. The methods chosen for comparison are certainly equivalence and stochastic control for unknown parameters without learning. The first employs the crude assumption that the unknown matrices A and C equal their point estimates $\bar{A}$ and $\bar{C}$ obtained from data up to time 0. The second has been described at the end of section 2 as a means of finding a tentative path to perform the quadratic approximation of (2.7). It takes into account the uncertainty of A and C as of the end of period 0, but does not anticipate future learning about them in the design of an optimal policy for the current period. One could very well revise his posterior density of A and C passively after period 1, but, according to the second method, he does not incorporate such revision in deriving the policy for period 1. The third method, that of the present paper, does take into account possible future learning in finding the current optimal policy. In the minimization at each future stage, the conditional expectation of the welfare loss utilizes the posterior density of A and C as of the time of the future decision, and not just at time 0 as in the second method.

It would be convenient to begin with the first method, that of certainty equivalence, assuming $K_{t,s}$ in the welfare function (2.2) to be zero for $t \neq s$. The solution is a simple modification of the second method, as given by equation (2.10), (2.11), (2.12) and (2.13). If the unknown parameters A and C are reduced to $\bar{A}$ and $\bar{C}$, their point estimates (or the means of their posterior density) as of time 0, we simply replace the expectations of functions of A and C by the same functions of $\bar{A}$ and $\bar{C}$. Thus, following (2.11) to (2.13) the optimal control for each period is given by the feedback equation

$$\hat{x}_t = -(\bar{C}' H_t \bar{C})^{-1} [(\bar{C}' H_t \bar{A}) y_{t-1} + \bar{C}' h_t] \tag{3.1}$$

and the minimum welfare cost from period t on is given by

$$\min_{x_t} E_{t-1} W_t = \frac{1}{2} y'_{t-1} Q_{t-1} y_{t-1} + y'_{t-1} q_{t-1} + r_t \tag{3.2}$$

where

$$Q_{t-1} \equiv Q^t_{t-1,t-1} = \bar{A}' H_t \bar{A} - \bar{A}' H_t \bar{C} (\bar{C}' H_t \bar{C})^{-1} (\bar{C}' H_t \bar{A}) \tag{3.3}$$

$$q_{t-1} \equiv q^t_{t-1} = \bar{A}' - (\bar{A}' H_t \bar{C})(\bar{C}' H_t \bar{C})^{-1} \bar{C}' h_t$$

$$r_t = -\frac{1}{2} h'_t \bar{C} (\bar{C}' H_t \bar{C})^{-1} \bar{C}' h_t + \frac{1}{2} E_o u'_t H_t u_t$$

By (2.10), H_t, h_t and d_t are determined by the difference equations

$$(3.4)\qquad \begin{aligned} H_t &\equiv H^t_{t,t} = H^{t+1}_{t,t} + Q^{t+1}_{t,t} = K_t + Q_t \\ h_t &\equiv h^t_{t,t} = h^{t+1}_{t,t} + q^{t+1}_{t,t} = k_t + q_t \\ d_t &= d_{t+1} + r_{t+1} \end{aligned} \quad .$$

In (3.4), we have utilized

$$(3.5)\qquad \begin{aligned} H^{t+1}_{t,t} &= H^{t+2}_{t,t} + Q^{t+2}_{t,t} = H^{t+2}_{t,t} = \ldots = H^T_{t,t} = K_{t,t} \equiv K_t \\ h^{t+1}_t &= h^{t+2}_t + q^{t+2}_t = h^{t+2}_t = \ldots = h^T_t = k_t \end{aligned}$$

since, by (2.13), $Q^t_{ij} = 0$ and $q^t_i = 0$ for $i < t-1$.

For each period t , the optimal policy $\hat{x}_t$ minimizes the expected value of a quadratic function in y_t , with H_t and h_t as coefficients. By (3.4), this quadratic function can be decomposed into two parts:

$$(3.6)\qquad \left(\tfrac{1}{2}\, y_t' K_t y_t + y_t' k_t\right) + \left(\tfrac{1}{2}\, y_t' Q_t y_t + y_t' q_t\right) .$$

The first part is the contribution of y_t to welfare loss as y_t appears directly in the welfare function. The second part is the minimum future welfare cost due to $y_{t+1}, y_{t+2}, \ldots, y_T$ assuming the future $x_{t+1}, \ldots, x_T$ to be optimally chosen; it is also a quadratic function of the initial condition y_t on which future decisions will have to be built. This decomposition shows the relative importance of setting y_t for direct contribution to welfare and for foundation building for the future. The optimal x_t is chosen to minimize the sum of the expectations of these two quadratic functions in y_t, one for current benefits and the other for future benefits.

The same decomposition and interpretation apply to the second method. The only difference is that in computing the quadratic function for optimal future cost, different coefficients Q_t and q_t will be used. They are given by (2.13) rather than (3.3). When uncertainty in A and C as of time 0 is allowed for, the minimum future welfare cost will have to be computed differently from the case of certain A and C. The difference between these two functions is explained in detail in Chow (1973b). For the present purpose, the comparison of Q_t and q_t as between (2.13) and (3.3) shows how the above uncertainty affects the weights given to the preparation for future optimization in the determination of the current policy.

The possibility of learning as treated in this paper does not invalidate the present-future decomposition of the quadratic function of y_t whose expected value is to be minimized by x_t. However, it makes the future component more complicated. When learning is absent, the only concern for y_t from the future point of view is that it affects the minimum expected future cost from $t+1$ on; it does not affect minimum expected cost from $t+2$ on because, the model (2.1) being first-order, the latter is a function of y_{t+1} alone and not of y_t.[5] When learning is present, y_t affects the optimal expected cummulative costs from all future periods on, for equation (2.10) implies

$$(3.7) \qquad H_t \equiv H_{t,t}^{t} = H_{t,t}^{t+1} + Q_{t,t}^{t+1} = K_{t,t} + Q_{t,t}^{t+1} + (Q_{t,t}^{t+2} + \dots + Q_{t,t}^{T})$$

and similarly for h_t. The terms in parentheses above show respectively the effects of y_t on the optimal expected future costs from $t+2$ on, from $t+3$ on, etc. These terms are treated as zero by the first two methods. They are actually non-zero because, in spite of the first order system, y_t affects not only y_{t+1} but the planner's conceptions of all future y's through its influence on his posterior densities of A and C in all future periods. Note that $Q_{t,t}^{t+1} \equiv Q_t$ in (3.7) also differs from the corresponding estimates given by the first two methods as it incorporates the effect of y_t on the posterior density of A and C in $t+1$ while the others do not.

Thus by comparing the Q_t of method two with $\sum_{i=t+1}^{T} Q_{t,t}^{i}$ of method three, one can measure the effect of learning on the weight given to the future-component of the quadratic function of y_t whose expectation is to be minimized. Presumably, when learning is allowed for, the future-component will receive more weight, as the above sum will measure. Therefore, besides providing a solution, the method of this paper gives an explicit measure of the effect of learning on the quadratic function to be minimized in each period. The comparison of this function for the three methods will give more information about the impact of learning than a simple comparison of the optimal value for x_t .

The above decomposition analysis can also be applied to the evaluation of other suboptimal methods. I will briefly mention two other methods which also incorporate some element of learning. One is a simplified version of method three, obtained by omitting all $Q^t_{i,j}$ for $i \neq j$ in the quadratic approximation of optimal expected future cost from period t on and otherwise following method three. It will also compute the sum $\sum_{i=t+1}^{T} Q^i_{t,t}$ which can be compared with the corresponding sum of method three to ascertain what is missing in the simplification. Second, as it has been suggested in the literature, one may choose to account for learning for only $M < (T-t)$ future periods.[6] This simplification can be implemented in the framework of this paper by performing the minimizations backward in time up to $t+M$ by method two, thus obtaining $H^{t+M}_{i,j}$ and $h^{t+M}_{i,j}$, and then using the method of this paper to complete the remaining minimizations. It will yield a partial sum $\sum_{i=t+1}^{t+M} Q^i_{t,t}$ for comparison with the sum $\sum_{i=t+1}^{T} Q^i_{t,t}$ of method three.

4. Some Numerical Results

In order to illustrate the use of the method of this paper and to measure the effect of learning on the decomposition between immediate and future welfare and on the optimal control policy, a simple model with a scalar dependent variable y_t^* is used. The explanatory variables are y_{t-1}^* and a scalar control variable x_t. To include both y_t^* and x_t in a vector y_t of dependent variables, we write

$$(4.1) \qquad \begin{bmatrix} y_t^* \\ x_t \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1}^* \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} c \\ 1 \end{bmatrix} x_t + \begin{bmatrix} u_t^* \\ 0 \end{bmatrix}$$

or, in the notation of (2.1),

$$(4.2) \qquad y_t = A\, y_{t-1} + C\, x_t + u_t \, .$$

Let n observations on the system (4.1) be available by the end of time 0; assume y_t and its explanatory variables to have been observed for $t = -n+1,\ldots,0$. Let the prior density of the parameters a and c be diffuse as of the beginning of period $-n+1$.[7] Using the results of Chow (1973 b), one can evaluate the expectations required in the crucial function (2.7) as follows. Let $\hat{a}_{T-1}$ and $\hat{c}_{T-1}$ denote respectively the least squares estimates of a and c using data from period $-n+1$ to period $T-1$, and let s_{T-1} denote the sum of squared residuals of y_t from the least-squares fitted regression, also from $-n+1$ to $T-1$. Let h_{ij} be the i-j element of the matrix H, omitting the superscript and subscripts of H in equation (2.7). The required expectations in (2.7) are

$$(4.3) \qquad E_{T-1}a = \hat{a}_{T-1} \; ; \qquad E_{T-1}c = \hat{c}_{T-1}$$

and

$$(4.4) \qquad E_{T-1}A'HA = \begin{bmatrix} h_{11} E_{T-1}a^2 & 0 \\ 0 & 0 \end{bmatrix} \; ;$$

$$E_{T-1}C'HA = [h_{11} E_{T-1}ac + h_{12} \hat{a}_{T-1} \quad 0] \; ;$$

$$E_{T-1}C'HC = h_{11} E_{T-1}c^2 + 2 h_{12} \hat{c}_{T-1} + h_{22}$$

where

$$(4.5) \qquad \begin{bmatrix} E_{T-1}a^2 & E_{T-1}ac \\ E_{T-1}ac & E_{T-1}c^2 \end{bmatrix} = \begin{bmatrix} \hat{a}^2_{T-1} & \hat{a}_{T-1}\hat{c}_{T-1} \\ \hat{a}_{T-1}\hat{c}_{T-1} & \hat{c}^2_{T-1} \end{bmatrix} + \frac{s_{T-1}}{(n+T-1)-4} \begin{bmatrix} \sum\limits_{-n+1}^{T-1} y^{*2}_{i-1} & \sum\limits_{-n+1}^{T-1} y^*_{i-1}x_i \\ \sum\limits_{-n+1}^{T-1} y^*_{i-1}x_i & \sum\limits_{-n+1}^{T-1} x_i^2 \end{bmatrix}^{-1}$$

Equations (4.3), (4.4) and (4.5) are sufficient to specify the function (2.7) and, with t replacing T, the minimum expected future loss from any period t onward. Note that the last term of (2.7) $E_{T-1}u_T' H_{T,T}^T u_T = \text{tr } H_{T,T}^T E_{T-1} u_T u_T'$, though unknown, is not to be influenced by the choice of x_T and can therefore be regarded as a constant for the purpose of deriving the optimal policy. In the computations below, the sample covariance matrix of the regression residuals using the n available observations is used to represent $E_{t-1} u_t u_t'$ $(t=1,\ldots,T-1)$ in each of the three methods. Although this calculation is only approximate, it will not affect the comparisons between the methods.

In the following example, y_t^* of equation (4.1) is represented by annual gross national product in billions of current dollars, and x_t is represented by annual government purchases of goods and services in billions of current dollars. Annual observations of these variables from 1953 to 1972 constitute the sample of 20 observations available before planning begins. The equqtion explaining GNP by lagged GNP and government expenditures G can be interpreted as a reduced form equation from a structure consisting of an identity $GNP = C + I + G$, a consumption function explaining C by GNP and GNP_{-1}, an investment function explaining I by GNP and GNP_{-1}, although one may not wish to take this structure too seriously. For whatever it is worth, the regression using the 20 annual observations is (with standard errors in parentheses)[8]

$$(4.6) \qquad \hat{y}^*_t = \underset{(.072)}{.890}\, y^*_{t-1} + \underset{(.311)}{.779}\, x_t \qquad \begin{array}{l} R^2 = .998 \\ s^2 = 159.1 \\ DW = 2.109 \end{array}$$

Assume that, at the beginning of period 1 (1973), one utilizes the sample data and the model of equation (4.6) to steer GNP and government expenditures toward their target paths by applying one of the three optimal control policies. The target paths specify a 6% annual growth rate for GNP from its 1972 figure, and a 5% annual growth rate for government expenditures from its 1972 figure. The $K_{t,t}$ matrix is assumed to be a 2×2 identity matrix for all t, assigning equal cost to the squared deviation of each of the two variables from its target. The planning horizon T is assumed to be 10 years. The regression (4.6) shows that the coefficient of the control variable x_t is significant at not much better than 5%. There seems to be about the right amount of uncertainty in this model to make the example interesting. If the standard errors of the coefficients were much smaller, one might not be able to observe the effects of uncertainty and of learning. If they were much larger, the model would probably not be taken seriously for planning purpose. When method III (the method of this paper) is applied, the result of the certainty equivalence solution (method I) is used to provide the tentative path, including the random residual in the equation, although the result of method II could very well have been used instead. Table 1 presents some results from applying the three methods of optimal control to this example.

Following the present-future decomposition of the quadratic function of y_t whose expectation is to be minimized at each stage, as it was discussed in section 3, one observes that, for method I (certainty equivalence), the $H^t_{t,t}$ matrix is reaching a steady-state value as t decreases from 10 to 1. For period 10, $H^{10}_{10,10}$ is simply $K^{10}_{10,10}$, the 2×2 identity matrix, since there will be no future to speak of after period 10. Then some more weight is added to the square of the y^*_t variable as t decreases, or as the future becomes longer in duration, until the future component becomes .654. The conditions under which the matrix difference equations (3.3) and (3.4) for $H_t = H^t_{t,t}$ will have a steady state solution have been discussed in Chow (1972a) and Chow (1973c) and will not be repeated here. The present example illustrates a steady-state solution for $H^t_{t,t}$. There is no weight given to the square of x_t in the future component of the quadratic function because all future y's from t+1 on will not be dependent on x_t (y^*_{t+1} is a function of y^*_t and x_{t+1}, but not of x_t).

It was pointed out in Chow (1973b) that, in the case of control under uncertainty without learning, the matrix $Q_{t-1} = Q^t_{t-1,t-1}$ of (2.13), namely the matrix of the quadratic

TABLE 1

COMPARING THREE CONTROL METHODS:

(I) Certainty Equivalence
(II) Unknown Parameters Without Learning
(III) Unknown Parameters With Learning

Method / Result	I	II	III
$H^{10}_{10,10}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
G_{10}	[-.4317 0]	[-3898 0]	[-.4115 0]
g_{10}	1261	1181	1222
$\hat{x}_{10}$	417.1	419.3	417.5
$H^{9}_{9,9}$	$\begin{bmatrix} 1.4933 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.5381 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.7290 & -.1162 \\ -.1162 & 1.0132 \end{bmatrix}$
G_9	[-.5434 0]	[-.4896 0]	[-.5401 0]
g_9	1400	1302	1395
$\hat{x}_9$	402.2	403.2	403.5
$H^{5}_{5,5}$	$\begin{bmatrix} 1.654 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.8043 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 2.3316 & -.2745 \\ -.2745 & 1.0250 \end{bmatrix}$
G_5	[-.5726 0]	[-.5266 0]	[-.6206 0]
g_5	1155	1087	1225
$\hat{x}_5$	311.2	311.3	310.6
$H^{1}_{1,1}$	$\begin{bmatrix} 1.654 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.8063 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 4.4859 & -1.5685 \\ -1.5685 & 2.2303 \end{bmatrix}$
G_1	[-.5727 0]	[-.5269 0]	[-.5351 0]
g_1	916	862	875
$\hat{x}_1$	254.3	253.7	256.8

function giving the minimum expected future cost (2.12) from period t on, can be interpreted as the covariance matrix of residuals in the weighted regression of A on C. Similarly, in the certainty equivalence solution, the matrix Q_{t-1} of (3.3) can be interpreted as the covariance matrix of residuals in the weighted regression of $\bar{A}$ on $\bar{C}$. Insofar as A and C can be viewed as $\bar{A}$ and $\bar{C}$ plus random errors, the former covariance matrix may be expected, under not unusual circumstances, to be not smaller than the latter matrix, in the sense that their difference is a positive semidefinite matrix. The present example illustrates this point, the matrix $Q_{t-1} = H^{t-1}_{t-1,t-1} - K_{t-1,t-1} = H_{t-1} - I$ having a larger leading term for method II than for method I. To put it in another way, the introduction of uncertainty increases the weight for the future component of the quadratic function to be minimized. When learning is introduced, one finds the weight given to the future component further increased, as the $H^{t}_{t,t}$ matrices for Method III given in Table 1 show. One should care more about the future if he is allowed to learn.

Turning now to the feedback control equations $x_t = G_t y_{t-1} + g_t$, (where y_{t-1} consists of y^*_{t-1} and x_{t-1}) one notices in Table 1 that the coefficient of y^*_{t-1} (or lagged GNP) is smaller in absolute value for method II than for method I. This shows that when uncertainty exists, one tends to respond less to changing circumstances. A more thorough discussion of this point can be found in Chow (1973b), in terms of the size of the coefficients G_t in the weighted regression of A on C, as compared with the coefficients in the regression of $\bar{A}$ on $\bar{C}$ (variables without errors). The coefficient of y^*_{t-1} by method III is larger in absolute value than by method II, suggesting that, if learning is allowed, it may pay to pursue a more active policy. An active policy is also indicated by the intercept g_t in the control equation. The role of g_t can be seen by considering the simpler model $y^*_t = c\, x_t + u_t$. The optimal setting (the intercept) for the one-period problem of minimizing $E_o(y^*_1 - z)^2$, z being the target, is $\hat{x}_1 = (E_o c)\, z/[\text{var } c + (E_o c)^2]$. Here more uncertainty as measured by a larger var c will tend to reduce the intercept. Insofar as learning and uncertainty may have opposite effects on G_t , the relative magnitudes of these coefficients as between methods III and I are indeterminate. It is important to observe that, in spite of the noticeable differences in the reaction coefficients, the numerical values of the optimal $\hat{x}_t$ by the three methods are remarkably similar; these values are obtained, for comparison purposes, by applying the different optimal control equations to the same set of values of y^*_{t-1} as used in the tentative path for method III. Presumably, a smaller (negative) coefficient in the feedback control equation is partly compensated by a larger (positive) intercept g_t , the latter playing the role of steering the variables to targets after the feedback effect of the former coefficient has been allowed for.

At this point, a number of questions will certainly occur to the reader. How will the results of this example change if the degree of uncertainty is increased or reduced (partly by increasing or reducing the standard error of the regression), if the time horizon is lengthened or shortened, if the sample values of x_t contain more or less variation, if the parameters in the model or in the welfare function are different, and if the model is larger, containing a system of several equations? How will the welfare loss compare as among the three methods? On the theoretical level, one might ask how close the quadratic approximation of this paper is to the truly optimal solution, and how the method can be modified or generalized to deal with systems of reduced form equations which are derived from structural equations (thus are subject to certain non-linear restrictions on the coefficients), with non-linear systems of dynamic equations, or with non-quadratic welfare functions. The answers to these and other related questions will have to await further investigations.

This paper has set forth an approximate method of optimal control of linear systems under uncertainty which incorporates learning, and has provided a framework to measure the effects of uncertainty and of learning on the optimal control policies. The small numerical example given illustrates the above effects. It tentatively suggests that, while the effects of uncertainty and of learning have turned out to be in the directions expected from theoretical analysis, the solution in terms of the optimal value of the instrument has not been materially affected. Undoubtedly, by changing the example such as increasing the standard error of the regression, one can intensify the effects of uncertainty and of learning on the optimal setting of the instruments. Nevertheless, for a wide range of situations in which the degree of uncertainty (however measured) in the parameters is not too large, one might find that the certainty equivalence solution, expecially when applied in an open-loop fashion, will turn out to be a reasonable first approximation to the optimal. I leave this conjecture to further investigations.

REFERENCES

G. C. Chow, "Optimal Control of Linear Econometric Systems with Finite Time Horizon," International Economic Review, Vol 13, No. 1 (Feb. 1972), pp. 16-25. (a)

___________, "How Much Could be Gained by Optimal Stochastic Control Policies," Annals of Economic and Social Measurement, Vol. 1, No. 4 (October, 1972) pp. 391-406. (b)

___________,"Multiperiod Predictions from Stochastic Difference Equations by Bayesian Methods," Econometrica, Vol. 41, No. 1 (January, 1973) pp. 109-118. (a)

___________, "Effect of Uncertainty on Optimal Control Policies," International Economic Review, Vol. 14, No. 3 (October, 1973), pp. 631-645. (b)

___________, "Problems of Economic Policy from the Viewpoint of Optimal Control," American Economic Review, Vol. 63, No. 5 (December, 1973). (c)

E. C. MacRae, "Linear Decision with Experimentation," Annals of Economic and Social Measurement, Vol. 1, No. 4 (October, 1972), pp. 437-448.

E. C. Prescott, "The Multiperiod Control Problem Under Uncertainty," Econometrica, Vol. 40, No. 6 (November, 1972), pp. 1043-1058.

G. C. Rausser and J. W. Freebairn, "A Comparison of Approximate Adaptive Control Solutions to the U.S. Beef Trade Policy," Annals of Economic and Social Measurement, Vol. 3, No. 1 (January, 1974), forthcoming.

E. Tse, "Adaptive Dual Control Methods," Annals of Economic and Social Measurement, Vol. 3, No. 1 (January, 1974), forthcoming.

FOOTNOTES

*I am much indebted to Andrew Abel for extremely able research assistance, to Ray C. Fair and other members of the Econometric Research Program seminar at Princeton for valuable suggestions and discussions, and to the National Science Foundation for financial support through Grant GS32003X.

1. The references in the control literature are too numerous to cite. In the economics literature, Prescott (1972) deals with the problem of learning using a very simple model but provides no new method of solution; its results were computed by complete enumeration. MacRae (1972) and Tse (1974) provide interesting approximations to the optimal solution, and are highly recommended to the reader.

2. The conversion of a higher-order system into first order and the incorporation of control variables in the left-hand-side vector y_t to simplify derivations are explicitly shown in Chow (1972a) and Chow 1972b). See also the illustrative example of section 4 below.

3. The reader will be easily convinced that adding this term will not affect the derivation of the method below, except to make the algebra lengthier.

4. See also the example of section 4 below.

5. Of course, by y_t we mean possibly lagged endogenous variables if the system (2.1) has been converted to first-order from an originally higher-order system.

6. See Rausser and Freebairn (1974) for some calculations using this solution and other suboptimal adaptive control solutions as applied to the U.S. beef trade policy.

7. If the prior density is informative, one can easily modify the following formulae. See Chow (1973a) for the required modifications.

8. Data for GNP and government purchases of goods and services from 1964 to 1972 are from _Survey of Current Business_, July issues of 1973 back to 1968, and, before 1963, from _The National Income and Product Accounts of the United States, 1929-1965_ (U.S. Department of Commerce, 1966). GNP from 1952 to 1972 are 345, 365, 365, 398, 419, 441, 447, 484, 504, 520, 560, 591, 632, 685, 750, 794, 864, 930, 977, 1,056 and 1,155. Government purchases of goods and services from 1953 to 1972 are 81.6, 74.8, 74.2, 78.6, 86.1, 94.2, 97.0, 99.6, 107.6, 117.1, 122.5, 128.7, 137.0, 156.8, 180.1, 199.6, 210.0, 219.5, 234.3 and 255.0.