#28

# ECONOMIC SYSTEMS MARKED BY OBSOLESCENCE OF INFORMATION

By

Ronald Britto

Discussion Paper Number 28
November 1972

ECONOMIC SYSTEMS MARKED BY OBSOLESCENCE OF INFORMATION*

## I. Introduction

In recent years there has been a revival of interest in stochastic economic models whose basic variables are unobservable and must consequently be estimated by the economic agent as part of the optimizing process.[1] Friedman's well-known "permanent income" hypothesis of consumption behavior is a notable example. M. Nerlove [ 7 ] and L. Taylor [ 8 ] have successfuly used such models to obtain conditions yielding estimators for the unobserved variables in the form of a distributed lag function of past observations. Estimators with this property have been widely used in econometric work, making it a matter of some importance that a solid theoretical underpinning should be established.

My goals here are less specific than the derivation of the distributed lag function as an optimal estimator. While I shall use an economic model much like those of Nerlove and Taylor in that it is characterised by unobserved variables - to be called states of the world - I intend to study its properties with the aid of the apparatus of decision-theory and information theory. Thanks to the work of Blackwell and Girshick [ 1 ], Marschak [ 5 ] and Marschak and Miyasawa [ 6 ] there exist powerful theorems that can be applied for this purpose. It is worth mentioning here that one property of interest, the obsolescence of information, suggests why estimation by means of distributed lags with declining weights might be optimal in our framework.

The plan of the paper is as follows. Section II contains an example of the kind of economic situation that the model describes and lists some of its attractive features. The next section is devoted to a study of conditions related to the obsolescence of information. In section IV we present a simulatic

---

* I have had many extremely helpful discussions with my colleague Joseph Ostroy on the topics of this paper, but remain responsible for the views express- Mr. Rakesh Sarin has provided me with valuable research assistance in connection with the simulation studies reported on in sections IV and V.

[1] Marschak [ 4 ] has emphasized that the 2-step procedure of first estimating the parameter of interest and then taking an action based on this estimate is generally incorrect.

of the working of the model for a simple case with two states of the world and two observations. The problem investigated by means of this example is that of the number of periods it takes a decision-maker on the average to realise that the state of the world has changed. Section V discusses the long-run properties of the model and section VI the formation of expectations. Our findings are summarised in section VII.

## II *The Model and Its Scope*

Let us consider the example of a seller of some good who must determine the demand curve facing him in the current period. He has a series of past observations of prices and quantities sold, which he can - and will - use in some way to obtain an estimate of the amounts that he will be able to sell at different prices today. However there are two reasons why he does not know whether in fact the past observations have all come from the same demand curve. Firstly, the demand curve in any period may be one of several, depending upon what actions competitors have taken, the incomes of buyers, changes in tastes, etc. Although he can guess at the degree to which these forces have been operating, he does not know their effects with full certainty. Secondly, observations will not lie on the demand curve that generates them because of transient factors. Calling the various possible demand curves states of the world, we assume that the process generating them is a first-order Markov-process, known to the decision-maker. This means that the conditional probability of a particular state occurring in period t, given that some state occurred in period t-1, is independent of states occurring before t-1. The decision-maker also knows, we assume, the conditional probability distribution of observations, given the state of the world. The knowledge of these probability distributions, together with the history of past observations, is all he has to go by.

We hasten to remark that this model is very broad in scope and can be applied to a variety of problems. A second set of examples is obtained when states of the world are interpreted as equilibrium exchange rates (or prices of any sort) and the observations are the current surpluses in the balance of payments of a country (or number of unfilled orders). *The model will be* applicable in all situations in which it is desirable to distinguish the effects on observations of changes in the underlying economic structure from the effects due to transient factors. This distinguishes it from the models used to analyse inventory problems, problems of determining optimal internationa reserves etc., where the underlying economic structure is assumed to remain unchanged. Thus in all the literature on optimal international reserves the assumption is made that the equilibrium exchange rates do not change over the entire time-interval being considered, with all changes in the balance-of-payments surplus ascribed to random factors.

Attractive features of the model outlined above are:

(i) It presents a picture of the world that is more realistic than the alternatives used hitherto. The world as an economic decision-maker sees it is in a constant rtate of flux, where the past is an unreliable guide to the present, but is used because it is the only guide available. On the other hand, the world is not pure anarchy. Our assumption that the stochastic process is a Markov process is intended to capture some elements of stability of the underlying structure.

(ii) It allows the decision-maker to form expectations about the future that are much less rigid than a straightforward extrapolation of past observations would be. These expectations will be affected by past observations in a manner that depends on the transition probabilities.

(iii) Related to (i) above, the model suggests that the standard practice in economic theory of separating the analysis of equilibrium from that of the stability of this equilibrium is quite misleading. There is no such thing as static, deterministic equilibrium; instead we have a series of processes that never settle down. Furthermore, the distinction between short-run and long-run reactions disappears. This is a desirable result, for there is no way in fact through which a decision-maker can immediately distinguish short-run changes from long-run changes.

(iv) Work done by Cyert and deGroot on the use of the Bayesian approach to duopoly theory is very similar in spirit to our model and suggests that it may be unnecessary to construct models on an ad hoc basis for the investigation of different problems in economic theory.

To summarise, the model is designed to study the adaptive behavior of economic units as they respond to changes in the stochastic environment. Both time and uncertainty are accorded paramount importance in this 'vision' of the operations of an economic system.

## III. The Obsolescence of Information

### Notation

$z_i(t)$ $(i = 1,\ldots.,m)$ indicates the i'th state of the world in period t.

$x_j(t)$ $(j = 1,\ldots.n)$ indicates the j'th observation in period t.

$y_k(t,\beta)$ $(k = 1,\ldots.,(t-\beta)n)$ is the k'th member of the set of sequences of observations from the periods $\beta,\ldots.,t-1$, where $\beta$ is a positive integer $< t$.

$P(t)$ is the m x m Markov transition matrix with elements $p_{ik}(t) = p(z_k(t+1)|z_i(t))$ $(i,k = 1,\ldots.,m)$, where p( ) refers to the probability of the event in parenthesis.

$Q(t)$ is the m x n Markov matrix with elements $q_{ij}(t) = p(x_j(t)|z_i(t))$.

$a_i(t)$ is the i'th action available in period t $(i = 1,\ldots.,\ell)$.

$p_o$ is an m-component vector giving the a priori probability for the various states of the world in the initial period.

$T(t)$ is an $\ell$ x n payoff matrix whose elements $\tau_{ij}(t)$ show the payoff to the decision-maker in period t when action $a_i$ is taken and observation $x_j$ is made.

$\alpha$ is the discount factor per time period $(0<\alpha<1)$.

$J(t,t_1)$ is an n x n Markov matrix whose elements $j_{ij}(t,t_1) = p(x_j(t_1)|p(x_i(t))$, where $t_1$ is any positive integer $\neq t$.

### Assumptions

A1. The action in any period t must be taken before the observation $x_j(t)$ is available, although all past observations are known. The states of the world are never known.

A2. The matrix $P(t)$ is constant over time and written as $P$. $P$ is non-singular.

A3. The matrices $Q(t)$ and $T(t)$ are constant over time and indicated by $Q$ and $T$ respectively.

A4. The rank of $Q$ is min $(m,n)$.

A5. The conditional probabilities $x_j|z_i$ are independent of the actions taken in any period, as well as past observations and states of the world.

A6. The decision-maker maximizes the discounted sum of the expected value of the payoffs over all periods. Symbolically, his objective function is

$$\sum_{t=0}^{+\infty} \sum_{k=1}^{n^t} \sum_{j=1}^{n} \alpha^t \tau_{ij}(t) p(y_k(t,0)|x_j(t)) p(x_j(t)),$$

where by convention $p(y_k(0,0)|x_j(0)) = 1$ for all $k,j$.

We turn now to some elementary implications of the above definitions and assumptions.

1. The *a priori* probability vector for the staes of the world in period $t = p_o P^t$.

2. The *a priori* probability vector for the observations in period $t = p_o p^t Q$.

A7. The vector $p_o p^t$ is strictly positive for all $t$.

3. The vector $p_o P^t Q$ is positive for all $t$. (From A7 and A4.)

4. Because of A5 and the assumption that the $z_1(t)$ are a first-order Markov process the maximization of the objective function can be broken down into a series of independent maximizations, one for each period. In all that follows we shall accordingly focus on the decision to be made in some specific period alone.

Consider the situation facing our decision-maker in period t. He has a series of past observations $y_k(t,0)$ and in addition knows the a priori probabilities of observations in the current period. He will use the past observations to revise these prior probabilities in Bayesian fashion, and make his decision with the aid of the revised probability vector. Knowing the transition matrix P, he could also use the past observations to generate a revised probability distribution for observations in future periods, a topic we shall disucss in the next section. At the moment we are interested in the value of the information provided by the past set of observations.

Let us commence with the simplest case, where there is only one past observation $x_j(\beta)$. Since the decision-maker's current payoff depends only on $a_i(t)$ and $x_j(t)$, the matrix $J(t,\beta)$ will determine the value of an observation from the period $\beta$. For instance, if this matrix should happen to equal I, the unit matrix, or a permutation thereof, there is perfect information and all other observations are unnecessary. At the other extreme is the case where all rows of $J(t,\beta)$ are identical.[1]

Now $p(x_j(\beta)|x_i(t)) = p(x_j(\beta),\ x_i(t))/p(x_i(t))$

$$= \sum_{rs}\sum p(z_r(\beta),\ z_s(t))\ p(x_j(\beta)|z_r(\beta))\ \ p(x_i(t)|z_s(t))/p(x_i(t))$$

$$= \sum_{rs}\sum p(z_r(\beta))\ p(z_s(t)|z_r(\beta))\ p(x_j(\beta)|z_r(\beta))\ p(x_i(t)|z_s(t))/p(x_i(t)).$$

The numerator of this last expression is, for fixed value of i and j, a quadratic form in the variables $p(z_r(\beta))\ p(x_j(\beta)|z_r(\beta))$ and $p(x_i(t)|z_s(t)$, where r and s = 1,....,m. It can consequently be expressed as the i,j'th element of a certain matrix, which is the product of three matrices. Without

---

[1] See [5], pp. 199, 200.

presenting the derivation we proceed immediately to write the product matrix, which is

$$Q'(P^{t-\beta})'P_o(\beta)^*Q,$$

where $P_o(\beta)^*$ is the m x m diagonal matrix whose main diagonal is the vector $p_oP^{\beta}$, and the primes indicate the transposes of the corresponding matrices.

$p(x_i(t))$ is of course the i'th component of the vector $p_oP^tQ$. Let $Q_o(t)^*$ denote the n x n diagonal matrix whose main diagonal is this vector. $(Q_o(t)^*)^{-1}$ exists by A7. Then it can be verified that the matrix whose i,j'th element is $p(x_j(\beta)|x_i(t))$,

$$J(t,\beta) = (Q_o(t)^*)^{-1}Q'(P^{t-\beta})'P_o(\beta)^*Q, \quad \beta = 1,\ldots,t-1 \qquad (1)$$

The process of deriving the elements of the matrix $J(t,\gamma)$, where $\gamma$ is any positive integer $>t$, is similar and yields

$$J(t,\gamma) = (Q_o(t)^*)^{-1}Q'P_o(t)^*P^{\gamma-t}Q, \quad \gamma = t+1,\ldots. \qquad (2)$$

Suppose we wish to compare the value of an observation from period $\beta_1$ with one from some preceding period $\beta_2$, where $\beta_2 < \beta_1$. The two matrices of conditional probabilities,

$$J(t,\beta_1) = A(t)P_o(\beta_1)^*Q, \qquad (3)$$

and $J(t,\beta_2) = A(t)(P^{\beta_1-\beta_2})'P_o(\beta_2)^*Q, \qquad (4)$

where $A(t) = (Q_o(t)^*)^{-1}Q'(P^{t-\beta_1})'$.

The expected value to the decision-maker of an observation from any past period $\beta$,

$$v(t,\beta) = \max_i \sum_{jk}\sum T_{ij}(t)\ p(x_k(\beta)|p(x_j(t))\ p(x_j(t)). \qquad (5)$$

How does $v(t,\beta_1)$ compare with $v(t,\beta_2)$? It is known that, for all payoff matrices T,

$$v(t,\beta_1) \geq v(t,\beta_2) \qquad (6)$$

if and only if

$$J(t,\beta_2) = J(t,\beta_1)M_1, \tag{7}$$

where $M_1$ is some n x n Markov matrix.[1]

From (3) and (4), it is sufficient for (7) that for some Markov matrix $M_1$,

$$(P^{\beta_1-\beta_2})'P_o(\beta_2)^*Q = P_o(\beta_1)^*QM_1,$$

or $BQ = QM_1$, where $B = P_o(\beta_1)^*)^{-1}(P^{\beta_1-\beta_2})'P_o(\beta_2)^*$.

Note that $(P_o(\beta_1)^*)^{-1}$ exists because of A7

Lemma 1

B is a Markov matrix, i.e., its elements $b_{ij} \geq 0$ $(i,j, = 1....,m)$, and $Bu = u$, where u is the m-component vector $(1,....,1)$.

Proof: It is clear that B, the product of three non-negative matrices, is non-negative.

Now $uP_o(\beta_2)^* = p_oP^{\beta_2}$.

Therefore $uP_o(\beta_2)^*P^{\beta_1-\beta_2} = p_oP^{\beta_2}P^{\beta_1-\beta_2} = p_oP^{\beta_1} = uP_o(\beta_1)^*$.

Therefore $uP_o(\beta_2)^*P^{\beta_1-\beta_2}(P_o(\beta_1)^*)^{-1} = u$.

Taking transposes, we obtain immediately

$$(P_o(\beta_1)^*)^{-1}(P^{\beta_1-\beta_2})'P_o(\beta_2)^*u = u$$

Let $\hat{Q} = \{q|\sum_i w_i q^i = q\}$, where $w_i \geq 0$ and $q^i$ is the i'th column of Q.

We are now ready to state

Theorem 1

Let $n \geq m$. A necessary condition that the value of information from period $\beta_1$ should be not less than the value of information from an earlier

---

[1] See [1], p. 328 (Theorem 12.2.2), or [6], p. 152 (Theorem 8.1).

period $\beta_2$, where $\beta_2 < \beta_1$, is: $Bq^i \in Q$ for all i (i=1,....,n).

Proof: Since $n \geq m$ and the rank of P is m, it follows from A4 that the rank of the matrix A(t) in (4) is m. $J(t,\beta_2) = J(t,\beta_1)M_1$ therefore implies that $QM_1 = BQ$, where $M_1$ is an n x n Markov matrix. Take the i'th column of BQ, $Bq^i$. This equals $Qm_1^i$, where $m_1^i$ is the i'th column of $M_1$. Since $m_1^i \geq 0$, $Bq^i \in \hat{Q}$ (i=1,....,n).

<u>Theorem 2</u>

Let $n \leq m$. Then a sufficient condition that the value of information from period $\beta_1$ should be not less than the value of information from an earlier period $\beta_2$, where $\beta_2 < \beta_1$, is: $Bq^i \in \hat{Q}$ for all i (i=1,....,n).

Proof: Since the rank of Q is n by A4, there exists an n x m matrix $Q^+$ such that $Q^+Q = I$. $Q^+ = (Q'Q)^{-1}Q'$. Also $Q^+u_m = Q^+Qu_n = Iu_n = u_n$, where $u_n$ ($u_m$) is the n-component (m-component) vector (1,....1) and we have used the fact that $Qu_n = u_m$, since Q is a Markov matrix. Now there exists by assumption a non-negative n x n matrix $M_1$ such that $QM_1 = BQ$. Also $BQu_n = Bu_m = u_m$. Therefore $Q^+BQu_n = Q^+u_m = u_n$. But $Q^+BQu_n = Q^+QM_1u_n = IM_1u_n$. Therefore $M_1u_n = u_n$ and $M_1$ is the required Markov matrix that satisfies $QM_1 = BQ$ and it follows easily that $J(t,\beta_2) = J(t,\beta_1)M_1$.

From Theorems 1 and 2 we obtain immediately

<u>Theorem 3</u>

Let m = n. Then a necessary and sufficient condition that the value of information from period $\beta_1$ should be not less than that from an earlier period $\beta_2$ is: $Bq^i \in \hat{Q}$ for all i (i = 1,....,m).

Suppose we wished to compare the Markov matrices whose i,j'th elements are $p(x_j(\beta_1)|z_i(t))$ and $p(x_j(\beta_2)|z_i(t))$ for the value of the information they provide. The two matrices, it can be easily shown, are $(P_o(t)^*)^{-1}(P^{t-\beta_1})'P_o(\beta_1)^*Q$ and $(P_o(t)^*)^{-1}(P^{t-\beta_2})'P_o(\beta_2)^*Q$ respectively, and we shall indicate them by

$K(t,\beta_1)$ and $K(t,\beta_2)$. The proof of the following theorem is omitted because it parallels that of Theorems 1 and 2.

<u>Theorem 4</u>

(i) A necessary condition that $K(t,\beta_2) = K(t,\beta_1)M_2$ for some n x n Markov matrix $M_2$ is: $Bq^i \in \hat{Q}$ for all i $(i=1,\ldots,n)$.

(ii) This condition is sufficient if $n \leq m$.

We proceed now to the general case. Let $L(t,\beta_1)$ be the $n \times n^{t-\beta_1}$ Markov matrix whose i,j'th element is $p(y_j(t,\beta_1)|x_i(T))$, where $y_j(t,\beta_1)$, to repeat, is a particular member of the set consisting of all possible sequences of observations from the preceding $t-\beta_1$ periods. Let $\hat{y}_j(\beta_1,\beta_2)$ represent the j'th member from the set consisting of past observations from the preceding $t-\beta_1-1$ periods plus the observation from the $\beta_2$'th period, where $\beta_2 < \beta_1$. $\hat{y}_j(\beta_1,\beta_2)$ is thus obtained from $y_j(t,\beta_1)$ by replacing the $\beta_1$'th observation by the $\beta_2$'th one. Let $\hat{L}(t,\beta_1,\beta_2)$ be the $n \times n^{t-\beta_1}$ Markov matrix whose i,j'th element is $p(\hat{y}_j(\beta_1,\beta_2)|x_i(t))$.

<u>Lemma 2</u>

If $K(t,\beta_2) = K(t,\beta_1)M_2$ for some Markov matrix $M_2$, then $K(\beta_o,\beta_2) = K(\beta_o,\beta_1)M_2$, where $\beta_1 < \beta_o < t$.

Proof: By assumption $(P_o(t)^*)^{-1}(P^{t-\beta_2})'P_o(\beta_2)^*Q = (P_o(t)^*)^{-1}(P^{t-\beta_1})'P_o(\beta_1)^*QM_2$.

Premultiplying both sides by the matrix $(P_o(\beta_o)^*)^{-1}((P^{-1})^{t-\beta_o})'P_o(t)^*$ we obtain $(P_o(\beta_o)^*)^{-1}(P^{\beta_o-\beta_2})'P_o(\beta_2)^*Q = (P_o(\beta_o)^*)^{-1}(P^{\beta_o-\beta_1})'P_o(\beta_1)^*QM_2$.

<u>Theorem 5</u>

If $K(t,\beta_2) = K(t,\beta_1)M_2$ for some Markov matrix $M_2$, then

(i) $J(t,\beta_2) = J(t,\beta_1)M_2$, and

(ii) $\hat{L}(t,\beta_1,\beta_2) = L(t,\beta_1)M_3$ for some $n \times n^{t-\beta_1}$ Markov matrix $M_3$.

Proof:

(i) By assumption

$$(P_o(t)^*)^{-1}(P^{t-\beta_2})'P_o(\beta_2)^*Q = (P_o(t)^*)^{-1}(P^{t-\beta_1})'P_o(\beta_1)^*QM_2.$$

Premultiplying both sides by the n x m Matrix $(Q_o(t)^*)^{-1}Q'P_o(t)^*$ we obtain the desired result.

(ii) The i,j'th elements of $L(t,\beta_1,\beta_2)$ and $L(t,\beta_1)$ are

$p(x_{j_{\beta_2}}(\beta_2),\ x_{j_{\beta_1+1}}(\beta_1+1),\ldots,\ x_{j_{t-1}}(t-1)\,|\,x_i(t))$ and

$p(x_{j_{\beta_1}}(\beta_1),\ldots,\ x_{j_{t-1}}(t-1)\,|\,x_i(t))$ respectively.

Now $p(x_{j_{\beta_2}}(\beta_2),\ x_{j_{\beta_1+1}}(\beta_1+1),\ldots,\ x_{j_{t-1}}(t-1)\,|\,x_i(t))$

$$= p(x_{j_{\beta_2}}(\beta_2),\ x_{j_{\beta_1+1}}(\beta_1+1),\ldots,\ x_{j_{t-1}}(t-1),\ x_i(t))/p(x_i(t))$$

$$= \Big[\Sigma_{r_{\beta_2}}\Sigma_{r_{\beta_1+1}}\ldots\Sigma_{r_t}\ p\{(x_{j_{\beta_2}}(\beta_2),\ x_{j_{\beta_1+1}}(\beta_1+1),\ldots,\ x_{j_{t-1}}(t-1),\ x_i(t)\,|$$
$$z_{r_{\beta_2}}(\beta_2),\ z_{r_{\beta_1+1}}(\beta_1+1),\ldots,\ z_{r_t}(t)\}$$
$$\times\ p(z_{r_{\beta_2}}(\beta_2),\ z_{r_{\beta_1+1}}(\beta_1+1),\ldots,\ z_{r_t}(t))\Big]/\ p(x_i(t)$$

$$= \Big[\Sigma_{r_{\beta_2}}\Sigma_{r_{\beta_1+1}}\ldots\Sigma_{r_t}\ p(x_{j_{\beta_2}}(\beta_2)\,|\,z_{r_{\beta_2}}(\beta_2))\ p(x_{j_{\beta_1+1}}(\beta_1+1)\,|\,z_{r_{\beta_1+1}}(\beta_1+1))\ldots$$
$$\ldots\ p(x_i(t)\,|\,z_{r_t}(t))\ p(z_{r_{\beta_2}}(\beta_2),\ z_{r_{\beta_1+1}}(\beta_1+1),\ldots,$$
$$z_{r_t}(t))\Big]\ /p|x_i(t))$$

$$= \Big[\Sigma_{r_{\beta_2}}\Sigma_{r_{\beta_1+1}}\ldots\Sigma_{r_t}\ p(x_{j_{\beta_2}}(\beta_2)\,|\,z_{r_{\beta_2}}(\beta_2))\ p(x_{j_{\beta_1+1}}(\beta_1+1)\,|\,z_{r_{\beta_1+1}}(\beta_1+1))\ldots$$
$$\ldots p(x_i(t)\,|\,z_{r_t}(t))\ p(z_{r_{\beta_2}}(\beta_2)\,|\,z_{r_{\beta_1+1}}(\beta_1+1))\ p(z_{r_{\beta_1+1}}(\beta_1+1)\,|$$
$$z_{r_{\beta_1+2}}(\beta_1+2))\ldots p(z_{r_{t-1}}(t-1)\,|\,z_{r_t}(t))\ p(z_{r_t}(t))\Big]/\ p(x_i(t))$$

$$= \left[\Sigma_{r_{\beta_1+1}} \ldots \Sigma_{r_t}\, p(x_{j_{\beta_2}}(\beta_2)|z_{r_{\beta_1+1}}(\beta_1+1))\{p(x_{j_{\beta_1+1}}(\beta_1+1)|z_{r_{\beta_1+1}}(\beta_1+1))\ldots \ldots p(x_i(t)|z_{r_t}(t))\; p(z_{r_{\beta_1+1}}(\beta_1+1)|z_{r_{\beta_1+2}}(\beta_1+2))\ldots p(z_{r_t}(t))\}\right]$$

$$/\; p|x_i(t))$$

Now, from Lemma 2

$$p(x_{j_{\beta_2}}(\beta_2)|z_{r_{\beta_1+1}}(\beta_1+1)) = \Sigma\, p(x_k(\beta_1)|z_{r_{\beta_1+1}}(\beta_1+1))m_{kj}\;,$$

where $m_{kj}$ is the $k_1j$'th element of the matrix $M_2$. Therefore, substituting in the last but one expression for $p(x_{j_{\beta_2}}(\beta_2)|z_{r_{\beta_1+1}}(\beta_1+1))$, we obtain

$$p(x_{j_{\beta_2}}(\beta_2),\, x_{j_{\beta_1+1}}(\beta_1+1), \ldots, x_{j_{t-1}}(t-1)|x_i(t))$$

$$= \left[\Sigma_{r_{\beta_1+1}} \ldots \Sigma_{r_t} \Sigma_k\, p(x_k(\beta_1)|z_{r_{\beta_1+1}}(\beta_1+1))m_{kj}\{\;\}\right]/p(x_i(t))$$

$$= \left[\Sigma_{r_{\beta_1}} \ldots \Sigma_{r_t}\Sigma_k\, p(x_k(\beta_1)|z_{r_{\beta_1}}(\beta_1))\; p(z_{r_{\beta_1}}(\beta_1)|z_{r_{\beta_1+1}}(\beta_1+1)m_{kj}\{\quad\}\right]$$

$$/\; p(x_i(t))$$

$$= \Sigma_k\{p(x_k(\beta_1),\, x_{j_{\beta_1+1}}(\beta_1+1), \ldots, x_{j_{t-1}}(t-1)|x_i(t)\}m_{kj}$$

Therefore, if the columns of the matrices $L(t,\beta_1)$ and $\hat{L}(t,\beta_1,\beta_2)$ are arranged so that each set of n columns contains in numerical order the set of n observations from periods $\beta_1$ and $\beta_2$ respectively, the matrix $M_3$ has the form

$$\begin{bmatrix} M_2 & 0 & \ldots\ldots & 0 \\ 0 & M_2 & \ldots\ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & & M_2 \end{bmatrix},$$

with $n^{t-\beta_1-1} M_2$ matrices along the principal diagonal and null matrices elsewhere. Since $M_2$ is a Markov matrix, $M_3$ will clearly be a Markov matrix as well.

From Theorems 4 and 5 we obtain immediately

<u>Theorem 6</u>

Let $n \leq m$ and $B_q^i \varepsilon \hat{Q}$, $i = 1,\ldots,n$.

Then $L(t,\beta_1,\beta_2) = L(t,\beta_1)M_3$ for some Markov matrix $M_3$.

Theorem 6 provides a sufficient condition that the value of the observations from the preceding $t - \beta_1$ periods should be not less than that from the preceding $t - \beta_1 - 1$ periods plus the observation from period $\beta_2$, where $\beta_2 < \beta_1$. This condition involves the matrices Q and B, the latter being $(P_o(\beta_1)^*)^{-1}(P^{\beta_1-\beta_2})'P_o(\beta_2)^*$. Since B (besides depending on P) depends not just on the difference $\beta_1 - \beta_2$ but on $\beta_1$ and $\beta_2$, it is desirable to have a condition free of the latter dependence.

We first note that, if $J(t,\beta_1)M_{\beta_1}$ for every $\beta_1 < t$, then $J(t,\beta_2) = J(t,\beta_1)M$ for some Markov matrix M for any $\beta_2<\beta_1$, a result implicit in the transitivity of the ordering by value of information. Thus it suffices to look at B for $\beta_2 = \beta_1 - 1$, i.e. $(P_o(\beta_1)^*)^{-1}P'P_o(\beta_1 - 1)^*$. If $(P_o(\beta_1)^*) \approx P_o(\beta_1-1)^*$, a condition that will obtain for Markov processes whose limiting probabilities exist when t is large, the elements on the principal diagonal of B will be approximately equal to those on the principal diagonal of P. Now let $B = \lambda I + (1-\lambda)C$, where $0\leq\lambda\leq 1$ and C is any m x m Markov matrix all of whose rows are equal.[1] Then $BQ = \lambda IQ + (1-\lambda)BC = \lambda Q + (1-\lambda)D$, where D is a Markov matrix with the same property as C. Thus the columns of D are scalar multiples of the vector $u_m = (1,\ldots,1)$. But $u_m \varepsilon \hat{Q}$ because $Qu_n = u_m$; therefore the columns of D, and hence those of $(1-\lambda)D\varepsilon\hat{Q}$. Thus $B_q^i\varepsilon\hat{Q}$ for all i. Taking $\lambda$ now close to 1, we find that B is a matrix with elements on the principal diagonal close to 1. Since these elements are approximately equal to elements

1. It is easy to show that any 2x2 matrix $B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$ can be written in this form provided $b_{11} > b_{21}$ and $b_{22} > b_{12}$.

on the principal diagonal of P, we have a condition on the latter matrix, viz. $P_{ii} \simeq 1 - \delta i$ ($\delta$ small), i=1,...,m, that should ensure that $B_q^i \varepsilon \hat{Q}$, i=1,..,n, independently of Q. In terms of our model, this condition means that the stochastic process governing the movement from one state to another is stable, the probability being high that the environment will stay in the same state from one period to the next.

It is also clear that, for n = m, the condition $B_q^i \varepsilon \hat{Q}$ (i = i,...,n) is likely to hold for Q≃I independently of B, since it is trivially satisfied in the limiting case where Q = I. Indeed, in this limiting case the only observation of value is that from the preceding period, for it serves to identify with probability 1 the state prevailing in that period. Any knowledge of states prevailing in earlier periods is of course redundant because of the Markov property of the stochastic process governing states of the world.

What is the significance of this obsolescence of information? As we mentioned at the outset, it suggests that the weights attached to different observations for the purpose of taking today's action will decline, the farther back from the past they come. Care must be taken in interpreting this statement, for it is not true for any individual decision-maker. The concept of the value of information applies to an entire information structure - hence we are asserting that the phenomenon of declining weights will occur for the average individual. This is to be contrasted with the Nerlove-Taylor approach, where the goal is to obtain estimates in the form of distributed lags for each individual.[1]

---

1. Because of their more stringent requirements, both Nerlove and Taylor placed restrictions on their payoff functions. In Nerlove's case this function took the form of minimising the mean-square-error. Taylor used a payoff function quadratic in the action and state variables. On the other hand Taylor allowed the payoff in each period to depend not just on the action of that period but on all actions taken in the past as well.

## IV. Lags in Adjustment to Changes in the Environment

We study in this section the case where n=m=2 and the matrix

$$P = \begin{bmatrix} p_1 & 1-p_1 \\ 0 & 1 \end{bmatrix}$$

State 2 is thus an absorbing state. We shall assume that the system has moved to state 2 in period 1 and that it was known by every individual that the system was in state 1 prior to that. Thus the prior subjective probability for state 1 in period 1 is $p_1$ for each individual. The question of interest is how long it takes the average individual to learn that the change has occurred, and how this depends on the matrix Q. It is to be expected that it will normally take a few periods before the revised probability for state 1 will have dropped to a reasonably low level (say .10) and that the greater the value of the information structure Q the shorter will be this lag.

For our example we put

$$p_1 = .9, \quad Q_1 = \begin{bmatrix} .8 & .2 \\ .1 & .9 \end{bmatrix} \quad \text{and } Q_2 = \begin{bmatrix} .6 & .4 \\ .3 & .7 \end{bmatrix}$$

It is easily verified that $Q_2$ is less informative than $Q_1$, i.e., that $Q_2=Q_1M$ for some Markov matrix M. With each of these information matrices the model was simulated for 20 runs, the length of a run being 20 periods. The results were as follows.

For each run a sequence of revised probabilities was obtained. The average of these revised probabilities in each period was then calculated. With the information structure of $Q_1$ the average revised probabilities for state 1 in periods 2, 3, 4 and 5, given the observations from the preceding periods, were .63, .27, .14 and .065 respectively. The average individual had revised his probability to below .10 after just three observations. The corresponding probabilities with $Q_2$ were .78, .63, .58 and .44. (For period 11 it was still as high as .19) When no observations are used at all the sequence of probabilities is of course $(.9)^2 = .81$, $(.9)^3 = .73$, $(.9)^4 = .66$ and $(.9)^5 = .59$.

The chief point we are trying to make by means of this simple example is that the introduction of uncertainty into models of economic behavior is sufficient to justify the use of lags <u>even when large numbers of participants are involved</u>. When there has been a change in the state of the system, it takes a few periods before this change is discerned by the average individual. The length of this lapse will be determined by the information pattern available to individuals.

## V. Long-run Properties of the Model

We look in this section at the long-run behavior of the revised probability vector for states in period t when t is large. The i'th component of this vector (when the j'th sequence of observations from the preceding t-1 periods has occurred) is $p(z_i(t)|y_j(t,1))$, $i = 1,..m$; $j = 1,\ldots,n^{t-1}$. We assume that the process described by the transition matrix P is ergodic and irreducible, so that $\lim_{t\to\infty} p_o P^t = p^*$ independently of $p_o$. Because of this, the dependence of the revised probability vector, which shall be indicated by $p_r(z(t))$, on $p_o$ will clearly diminish as t increases. The question of interest is whether $\lim_{t\to\infty} p_r(z(t))$ exists under these circumstances.

It is easy to see that the answer is generally in the negative. Assume that n=m and put Q = I, the identity matrix. Then for any t the state in period t-1 is known exactly by every individual. It follows that $p_r(z(t)) = e_i P$, if the system was in state i in period t-1, where $e_i$ is the i'th unit vector. If $p^*$ contains more than one positive component, the system will switch from one state to another and $p_r(z(t))$ will fluctuate accordingly. More specifically, for large t $p_r(z(t))$ will equal $e_i P$ with probability $p_i$. This is true for each individual and hence also holds for the average individual.

It appears that the behavior of $p_r(z(t))$ for large t will depend on the information matrix Q and one is tempted to conjecture that the fluctuations will be smaller in amplitude as the value of Q decreases, since it is clear that $\lim_{t\to+\infty} p_r(z(t)) = p^*$ if Q has zero information value. For the case where only one observation is used the above conjecture is easily validated. It was mentioned in section III that the matrix $K(t,\beta_1)$, whose i,j'th element is $p(x_j(\beta_1)\,|z_i(t))$, $= (P_o(t)^*)^{-1}(P^{t-\beta_1})'P_o(\beta_1)^*Q$. Let us take two information matrices $Q_1$ and $Q_2$ such that $Q_2 = Q_1M$ for some Markov matrix M. It follows easily that $K(t,\beta_1)_{Q_2} = K(t,\beta_1)_{Q_1}M$, where the subscript denotes the specific information matrix used to obtain $K(t\beta_1)$. Let $N(t,\beta_1)_Q$ be the nxm matrix whose i,j'th element is $p(z_i(t)|x_j(\beta_1)$, i=i,...n; j=1,...,m, where Q represents the information matrix. Then it follows from a theorem proved by Marschak and Miyasawa ([6], Th. 8.2, p. 154) that $N(t,\beta_1)_{Q_2} = M^*N(t,\beta_1)_{Q_1}$, where $M^*$ is some nxn Markov matrix. Since the rows of $N(t,\beta_1)_Q$ give the revised probabilities, this implies that the revised probabilities obtained with information structure $Q_2$ are convex linear combinations of those obtained with $Q_1$. Thus the set of revised probability vectors obtained with $Q_2$ in any period is a subset of that obtained with $Q_1$ for the same period. This means that the amplitude of the fluctuations of the revised probability vector will be greater as the value of the information matrix increases.

We conjecture that this result is also valid in the general case, where all past observations are used. To gain an idea of the extent of the fluctuations the model was simulated for 100 periods with the following matrices:

$$P = \begin{bmatrix} .8 & .2 \\ .3 & .7 \end{bmatrix} ; \; Q_1 = \begin{bmatrix} .8 & .2 \\ .1 & .9 \end{bmatrix}, \; Q_2 = \begin{bmatrix} .6 & .4 \\ .3 & .7 \end{bmatrix} .$$

In each case the system started in state 2. When $Q_1$ was used, the revised probability of state 1 fluctuated between .783 and .354 On the other hand, the fluctuations ranged between .716 and .471 when $Q_2$ was used. Note that $p^* = (.6,.4)$ and that therefore, were Q equal to I, the revised probability would be .8 roughly 60% of the time and .3 the remaining 40% of the time.

What are the implications of this result? The most important one appears to be the following. If the action of a decision-maker is period t depends on $p_r(z(t))$ alone, as would be the case with the assumption about the payoff matrix made in section III, one would expect more frequent changes in the optimal action as the matrix Q becomes more informative, not merely for the individual decision-maker, but also for the group as a whole. To put the matter in a form that makes it seem perfectly reasonable, economic responses will tend to follow fluctuations in the underlying structure more closely if the pattern of information available to decision-makers is good and will remain more or less constant in the opposite case. An interesting possibility emerges as a corollary: if a public agency attempts to improve the flow of information to decision-makers, this might have the unexpected effect of increasing the fluctuations of the economic variables that the latter control![1]

---

1. Of course our model presupposes that the environment itself is subject to shifts. Also to be stressed is the hidden assumption that the Markov process described by the matrix P is unaffected by the actions of the individuals involved. In a full-fledged general equilibrium model P itself would be endogenous.

## VI The Formation of Expectations

We turn now to the revision of the probability distribution of observations in future periods as a result of past and current observations. The theorems here are essentially similar to those of ~~the preceding~~ section III and we shall therefore dispense with proofs in general. If $K(t,\gamma)$ is the $m \times n$ Markov matrix whose i,j'th element is $p(x_j(\gamma)|z_i(t))$, where $\gamma > t$, then

$$K(t,\gamma) = P^{\gamma - t}Q, \qquad \gamma = t+1, t+2, \ldots..$$

Similar to Theorem 4 is

**Theorem ~~6~~ 7**

Let $\gamma_2 > \gamma_1 > t$.

(i) A necessary condition that $K(t,\gamma_2) = K(t,\gamma_1)M_4$ for some Markov matrix $M_4$ is:

$$P^{\gamma_2 - \gamma_1} q^i \in \hat{Q} \text{ for all } i \quad (i = 1,\ldots., n).$$

(ii) If $n \leq m$, a sufficient condition that

$K(t,\gamma_2) = K(t,\gamma_1)M_4$ for some Markov matrix $M_4$ is: $Pq^i \in \hat{Q}$ for all i.

Let $F(\beta,\gamma)$ be the ~~(t-β)~~ $n^{t-\beta} \times n$ Markov matrix whose i,j'th element is $p(x_j(\gamma)|y_i(t,\beta))$, i.e. the conditional probability of observing $x_j$ in period $\gamma$ (where $\gamma > t$), given the particular sequence i of observations from the preceding $t-\beta$ periods. A theorem analogous to Theorem 5 (ii) is

**Theorem ~~7~~ 8**

If $K(t,\gamma_2) = k(t,\gamma_1)M_4$ for some Markov matrix $M_4$, where $\gamma_2 > \gamma_1$, then

$$F(\beta,\gamma_2) = F(\beta,\gamma_1)M_4.$$

What are the implications of Theorem 7? Let $f_i(\beta,\gamma)$ be the i'th row of $F(\beta,\gamma)$. Let x be a random variable. Then $f_i(\beta,\gamma)\, x$, where x is the vector $x = (x_1,\ldots.,x_n)$ of observations, is the expected value of the random variable x in period $\gamma$, given the sequence of observations $y_i(t,\beta)$. Let $\bar{x} = \max\,(x_1,\ldots.,x_n)$

and $\underline{x} = \min (x_1 \dots, x_n)$. If $F(\beta,\gamma_2) = F(\beta,\gamma_1)M_4$, then $F(\beta,\gamma_2)\ x = F(\beta,\gamma_1)M_4x$. Since $M_4$ is a Markov matrix, any component of the vector $M_4x$ is a weighted average of the components of x and lies therefore between $\underline{x}$ and $\bar{x}$ . Thus the range of the conditional expected value of x in period $\gamma_2$ is not greater than its range in period $\gamma_1$.

## VII Summary

The model studied in this paper is essentially the same as those used in earlier studies by Nerlove [7] and Taylor [8] to justify the use of distributed lags in econometrics. The unknown states of the world are assumed to be governed by a known Markov process; information about these states is provided by a set of observations more or less closely related to the states. The matrix whose typical element is the conditional probability of a particular observation, given a particular state, is called the information matrix.

Our most important result, discussed in section III, bears on the obsolescence of information, meaning that knowledge of an observation is less valuable on the average for current decision-making, the farther back in the past it lies. Conditions are given in terms of the matrix of transition probabilities and the information matrix that yield this result for the case where the number of observations is less than or equal to that of states. The value of the result lies in the suggestion that the model can justify the use of distributed lags with declining weights for any payoff functions, provided one is looking at the behavior of the typical individual.

In section IV we argue that the uncertainty about states of the world is responsible for lags in adjustment to a shift to a new state, the length of the lag depending on the information structure, and demonstrate this by some simple examples. The long-run revised probability vectors for states of the world are shown to fluctuate in section V, even when the prior probabilities converge to some limiting values, the extent of these fluctuations again depending on the information matrix. Finally, in section VI the same considerations giving rise to the obsolescence of information are shown to imply that the elasticity of expectations declines, the farther out in the future they lie.

# REFERENCES

[1] Blackwell, D. and M. A. Girshick, Theory of Games and Statistical Decisions. New York: John Wiley & Sons, 1954.

[2] Cyert, R. M. and M. H. De Groot, "Multiperiod Decision Models with Alternating Choice as a Solution to the Dupoly Problem", Quarterly Journal of Economics, August 1970, vol. 84(3), pp. 410-29.

[3] ________, "Bayesian Analysis and Dupoly Theory", Journal of Political Economy, Sept.-Oct. 1970, vol. 78(5), pp. 1168-84.

[4] Marschak, J., "On Adaptive Programming", Management Science, July 1963, vol. 9(4), pp. 517-26.

[5] ________, "Economics of Information Systems", Journal of the American Statistical Association, March 1971, vol. 66 (333), pp. 192-219.

[6] Marschak, J. and K. Miyasawa, "Economic Comparability of Information Systems", International Economic Review, June 1968, vol. 9(2), pp. 137-74.

[7] Nerlove, M., "Distributed Lags and Unobserved Components in Economic Time Series", in Ten Economic Studies in the Tradition of Irving Fisher. New York: John Wiley & Sons, 1967.

[8] Taylor, L., "The Existence of Optimal Distributed Lags", Review of Economic Studies, Jan. 1970, vol. 37 (1), pp. 95-106.