INTERNATIONAL ECONOMIC REVIEW
Vol. 30, No. 3, August 1989

OPTIMAL CONTROL OF AN UNKNOWN LINEAR PROCESS WITH LEARNING

BY NICHOLAS M. KIEFER AND YAW NYARKO[1]

Optimal control of a linear process with unknown parameters is considered when the horizon is infinite and rewards are discounted. Active learning strategies are considered, i.e., agents consider the information value of possible actions as well as current reward. Distributional assumptions are minimal in that no restriction to conjugate families is made. Convergence of beliefs and actions is established.

1. INTRODUCTION

Rational expectations models raise naturally the question of learning. The hypothesis that agents process information efficiently in an effort to learn about their environment is much stronger than the hypothesis that profit opportunities do not systematically go unexploited. However, the efficient information-processing hypothesis is itself much weaker than the hypothesis that agents actively seek to learn about their environment, even when learning is costly. This is the case known as "active learning." A natural modelling strategy is to assume that agents, faced with a tradeoff between current-period reward and information generation, allocate their efforts optimally given their beliefs about the economy. This turns out to be a difficult problem to study, and we focus attention in this paper on the optimizing behavior of an agent in an economy in which his behavior can generate information.

For simplicity the process that we study is the linear regression process with independent errors. The agent expresses his beliefs about unknown parameters, which can include parameters of the "error" process as well as regression coefficients, in the form of a probability distribution. At date t the agent chooses an action on the basis of his beliefs at that date. The action is chosen taking account of the one-period reward resulting from the action and of the value of the expected information gain from that action. After the action is taken an outcome is observed. The outcome, together with the action, add to the agent's stock of information about the process generating the data. We assume that the agent processes this information in accordance with the laws of probability, i.e., by Bayes' Rule. In this paper we are concerned with the existence of the optimal strategy, with the limiting behavior of the sequence of the agent's beliefs, and with the limiting behavior of the optimal action.

The linear structure has been used before in economics to study the question of

[1] We thank David Easley and Ingmar Prucha for helpful discussions on the topic treated here. Related material has been presented at the NBER-NSF Seminar on Bayesian Inference, the University of Iowa and the Conference on Dynamic Econometrics at Austin, Texas. We are grateful to participants for their suggestions. This research is supported in part by the NSF.

learning about parameters when the regression is subject to control by Taylor (1974), who studies (essentially) a simple linear regression model with known intercept and unknown slope; by Anderson and Taylor (1976) who study the problem with both slope and intercept unknown using a simulation experiment; and recently by Jordan (1985), who studies analytically the problem with unknown slope and intercept and gives conditions under which learning occurs (in the sense that parameter ''estimates'' are consistent). All of these papers consider the least-squares certainty-equivalence control rule with sequential updating. With this strategy least squares estimates of the parameters, based on past data, are substituted for true values in the one-period reward function and the action maximizing this reward function is taken. Then, when the outcome is observed the least-squares estimates are updated. A potential drawback of the least squares control rule is the substitution of least-squares estimates for true parameter values. It is natural to suspect that a simple improvement over this rule can be obtained by substituting the action which maximizes expected reward, where the expectation is taken over a distribution for the parameters, for the action which maximizes reward conditional on the least-squares estimates. For the purpose of comparing strategies, one might interpret the sampling distribution of the least-squares estimates as the appropriate distribution. This strategy is suggested by Zellner (1971). Harkema (1975) compares the certainty-equivalence and minimum expected loss rules and finds the latter superior in a problem with quadratic loss. Of course, neither of these strategies is likely to be optimal when rules which take account the information value of actions are considered. Other related work includes Chow (1981) and Prescott (1972).

Both of the strategies described allow only ''passive'' learning, so they might be expected to be inferior to a strategy which allows for ''active'' learning, i.e., a strategy which considers the information value of an action at each date. Our approach studies the ''full'' optimization problem, taking account of the information value of actions. Thus our policies, though difficult to calculate must do at least as well as the least-squares rule, and in view of Harkema's results can be expected to do better. One would expect that considerations involving information relevant to future decisions will play an important role when current information is sketchy and the future is not heavily discounted. These comments are supported by the example of Section 6 and by the related analysis of Easley and Kiefer (1988).

The formal analysis proceeds as follows. Section 2 sets up the general framework, illustrating the definitions with the ''normal-normal'' example. We stress that the results of this paper make no use of specific distributional assumptions or conjugate families, the ''normal-normal'' example is used only to illustrate the definitions and results. We define partial histories as sequences of actions, outcomes and beliefs; admissible partial histories are partial histories in which the sequence of beliefs evolve according to Bayes' Rule. We make assumptions on the utility function and define the reward (expected utility) function. The value function is defined, along with the optimal policy. Section 3 shows that stationary optimal policies exist, and that the value function exists and satisfies the usual functional equation. In Section 4 we consider the limiting behavior of the sequence of posterior distributions. We establish, using a Martingale convergence argument the

weak convergence of the sequence of beliefs. The limit distribution need not be point mass at the true values and it need not be centered at the true parameter values. Simplifying to the simple (one x variable) regression process, we obtain some properties of the class of posterior distributions which can arise as limits. In Section 5 we establish the result that the sequence of actions converges, and that the limit action is the optimal one-period action for the limit beliefs. Section 6 specializes to a "normal-normal" example, and illustrates that limit beliefs need not be centered at true values. This case arises for sufficiently low, but nonzero, discount factors. Proofs are primarily given in the Appendix.

Before turning to the formal analysis, we note that our results concern the optimizing behavior of an economic agent acting in an uncertain environment when there is a possibility of taking actions in order to generate information. We have seen that there is a trade-off between immediate reward and information accumulation. Implications for the amount of information revealed in economic equilibrium are not pursued here. Equilibrium with agents "learning" by using least-squares estimates of linear models has been studied recently by Anderson and Sonnenschein (1985). That paper establishes the existence of a rational expectations equilibrium but abstracts from dynamics and from the question of active versus passive learning on the part of agents in the economy. Blume and Easley (1984) study a dynamic equilibrium model in which agents process exogenous and predetermined data optimally. They find that the market process generates data which allow consistent estimation of the parameters of their economy when the parameter set is discrete. That paper contains further references to the literature on "revealing rational expectations equilibria." It seems a sensible research strategy to try to incorporate active learning into an equilibrium model.

2. THE MODEL: STRUCTURE, UNCERTAINTY, POLICIES AND REWARDS

In this section we sketch the general framework we wish to study and establish a specific example, the normal simple regression model with known variance and a normal conjugate prior distribution, which we will carry along as a particular application and which we will treat in Section 6. The example is closely related to that studied by Anderson and Taylor (1976), and Jordan (1985). Again we stress that the example is used to illustrate the terminology and results, and is not required for any of the main results.

Let Ω' be a complete and separable metric space, let F' be its Borel field, and (Ω', F', P') a probability space. Define on (Ω', F', P') the stochastic process $\{\varepsilon_t\}_1^\infty$, the shock process, which is unobserved by the agent. The shock process is assumed to be independent and identically distributed, with the common marginal distribution $p(\varepsilon_t \mid \phi)$ depending on some parameter, ϕ in R^h, which is unknown to the agent. We assume that the set of probability measures, $\{p(\cdot \mid \phi)\}$, is continuous in the parameter ϕ (in the weak topology of measures); and that for any ϕ, $\int \varepsilon p(d\varepsilon \mid \phi) = 0$. Let $\bar{X}$, the *action space*, be a compact subset of R^k. Define $\Theta = R^1 \times R^k \times R^h$ to be the parameter space. If the "true parameter" is $\theta = (\alpha, \beta, \phi) \in \Theta$, and the agent chooses an action $x_t \in \bar{X}$ at date t, then the agent observes y_t, where,

$$y_t = \alpha + \beta x_t + \varepsilon_t \tag{2.1}$$

and where ε_t is chosen according to $p(\cdot \mid \phi)$.

Our example is the simple linear regression model with unknown slope and intercept and with the ε_t independent draws from the standard normal distribution. In our example Ω' is R^∞, F' is the collection of Borel sets on R^∞, and P' is the infinite product of independent univariate normal disitribution with mean zero and variance one. There is no unknown parameter ϕ of the distribution of ε, but an example of the more general case could be obtained by letting the variance of ε be unknown. The action space $\bar{X}$ in the example is a closed interval in R^1. The parameter space Θ is R^2, corresponding to the unknown slope and intercept in the linear equation generating y:

$$y_t = \alpha + \beta x_t + \varepsilon_t.$$

Let $\tilde{\Theta}$ be the Borel field of Θ, and let $P(\Theta)$ be the set of all probability measures on $(\Theta, \tilde{\Theta})$. Endow $P(\Theta)$ with its weak topology, and note that $P(\Theta)$ is then a complete and separable metric space (see, e.g., Parthasarathy 1967, Ch. II, Theorems 6.2 and 6.5). Let $\mu_0 \in P(\Theta)$ be the prior probability on the parameter space, with finite first moment.

The agent is assumed to use Bayes' rule to update the prior probability at each date after any observation of (x_t, y_t). For example, in the initial period, date 1, the prior distribution is updated after the agent chooses an action x_1, and observes the value of y_1. The updated prior, i.e., the posterior, is then $\mu_1 = \Gamma(x_1, y_1, \mu_0)$, where $\Gamma: \bar{X} \times R^1 \times P(\Theta) \to P(\Theta)$ represents the Bayes' rule operator. If the prior, μ_0, has a density function, then the posterior may be easily computed. In general, the Bayes' rule operator may be defined by appealing to the existence of certain conditional probabilities (see the Appendix). Under standard conditions the operator Γ is continuous in its arguments, and we assume this throughout. Any $\{x_t, y_t\}$ process will therefore result in a posterior process, $\{\mu_t\}$, where for all $t = 1, 2, \ldots,$

$$\mu_t = \Gamma(x_t, y_t, \mu_{t-1}) \tag{2.2}$$

The prior probability for our example is the bivariate normal distribution for α and β,

$$p(\alpha, \beta \mid \bar{\alpha}_0, \bar{\beta}_0, \Sigma_0) = n\left(\begin{pmatrix}\bar{\alpha}\\ \bar{\beta}\end{pmatrix}_0, \Sigma_0\right).$$

This is the conjugate prior for our model, so all subsequent distributions μ_t will also be normal distributions. We index these distributions by their parameters $(\alpha_t, \beta_t, \Sigma_t)$ and note that the updating rule corresponding to (2.2) is (with $X_t = (1, x_t)$)

$$\Sigma_{t+1} = \Sigma_t - \Sigma_t X_t'(I + X_t \Sigma_t X_t')^{-1} X_t \Sigma_t$$

$$\begin{pmatrix}\bar{\alpha}\\ \bar{\beta}\end{pmatrix}_{t+1} = \Sigma_{t+1}[X_t' y_t + \Sigma_t^{-1}(\bar{\alpha}, \bar{\beta})_t]$$

see, e.g., Zellner (1971, Section 3.2.3).

Let $\bar{H}_n = P(\Theta) \times \Pi_{i=1}^{n-1} [\bar{X} \times R^1 \times P(\Theta)]$. A *partial history*, h_n, at date n is any element $h_n = (\mu_0, (x_1, y_1, \mu_1,) \ldots, (x_{n-1}, y_{n-1}, \mu_{n-1})) \in \bar{H}_n$; h_n is said to be

admissible if (2.2) holds for all $t = 1, 2, \ldots, n-1$. Let H_n be the subset of $\bar{H}_n$ consisting of all admissible partial histories at date n.

A policy is a sequence $\pi = \{\pi_t\}_{t=1}^{\infty}$, where for each $t \geq 1$, the policy function π_t: $H_t \to \bar{X}$ specifies the date t action $x_t = \pi_t(h_t)$, as a Borel function of the partial history, h_t in H_t, at that date. A policy function is stationary if $\pi_t(h_t) = g(\mu_t)$ for each t, where the function $g(\cdot)$ maps $P(\Theta)$ into $\bar{X}$.

Define $(\Omega, F, P) = (\Theta, \bar{\Theta}, \mu_0) \times (\Omega', F', P')$. Any policy, π, then generates a sequence of random variables $\{(x_t(\omega), y_t(\omega), \mu_t(\omega)\}_{t=1}^{\infty}$ on (Ω, F, P) as described above, using (2.1) and (2.2) (the technical details are stated in the Appendix).

For any $n = 1, 2, \ldots$, let F_n be the sub-field of F, generated by the random variables (h_n, x_n). Notice that x_n is F_n-measurable but y_n and μ_n are not F_n-measurable. Next define $F_\infty = V_{n=0}^{\infty} F_n$, the σ-algebra generated by $h_1, h_2, \ldots$.

We now discuss the utility and reward functions and define the optimality criterion. Let $u: \bar{X} \times R^1 \to R^1$ be the utility function, and, in particular, $u(x_t, y_t)$ is the utility to the agent when action x_t is chosen at date t and the observation y_t is made. We assume,

(A.1) u is uniformly bounded and continuous.

The reward function, $r: \bar{X} \times P(\Theta) \to R^1$, is defined by

$$r(x_t, \mu_{t-1}) = \int_{\Theta} \int_{R} u(x_t, y_t) p(d\varepsilon_t \mid \phi) u_{t-1}(d\alpha\ d\beta\ d\phi) \tag{2.3}$$

where $y_t = \alpha + \beta x_t + \varepsilon_t$.

Let δ in $(0, 1)$ be the discount factor. Any policy π generates a sum of expected discounted rewards equal to

$$V_\pi(\mu_0) = \int \sum_{t=1}^{\infty} \delta^{t-1} r(x_t(\omega), \mu_{t-1}(\omega)) P(d\omega) \tag{2.4}$$

where the (x_t, μ_t) processes are those obtained using the policy π. A policy π^* is said to be an *optimal policy* if for all policies π and all priors μ_0 in $P(\Theta)$,

$$V_{\pi^*}(\mu_0) \geqq V_\pi(\mu_0). \tag{2.5}$$

Even though the optimal policy, π^* (when it exists) may not be unique, the value function $V(\mu_0) = V_{\pi^*}(\mu_0)$ is always well-defined.

3. EXISTENCE OF A STATIONARY OPTIMAL POLICY

We now indicate that stationary optimal policies exist, and that the value function is continuous.

THEOREM 3.1. *A stationary optimal policy $g: P(\Theta) \to \bar{X}$ exists. The value function, V, is continuous on $P(\Theta)$, and the following functional equation holds:*

$$V(\mu_0) = \max_{x \in \bar{X}} r(x_1, \mu_0) + \delta \int V(\mu_1) p(d\varepsilon_1 \mid \phi) \mu_0(d\alpha\ d\beta\ d\phi) \tag{3.1}$$

where $\mu_1 = \Gamma(x, y_1, \mu_0)$ *and* $y_1 = \alpha + \beta x + \beta x + \varepsilon_1$, *and where the integral is taken over* $\Theta \times R^1$.

PROOF. Let $S = \{f: P(\Theta) \to R \mid f \text{ is continuous and bounded}\}$. Define $T: S \to S$ by

$$Tw(\mu) = \operatorname*{Max}_{x \in \bar{X}} \{r(x, \mu) + \delta \int w(\mu_1) p(d\varepsilon_1 \mid \phi) \mu(d\alpha\ d\beta\ d\phi)\} \tag{3.2}$$

where $\mu_1 = \Gamma(x, y_1, \mu)$ and $y_1 = \alpha + \beta x + \varepsilon_1$. One can easily show that for $w \in S$, $Tw \in S$; and that T is a contraction mapping. Hence there exists a $v \in S$ such that $v = Tv$. Replacing w with v in (3.2) then results in (3.1); and since $v \in S$, v is continuous. Finally, it is immediate that the solution to the maximization exercise in (3.2) (replacing w with v) results in a stationary optimal policy function (see Blackwell 1965, or Maitra 1968 for the details of the above arguments). Q.E.D.

We assumed in condition (A.1) that the utility function u was uniformly bounded. If we relax this and require only that u be bounded above then Theorem 3.1 above still holds except that we can now assert only the upper semicontinuity of the value function. (And, of course, we have to consider the possibility that some or all policies from an initial prior may yield total discounted return of $-\infty$.) Consult Schal (1975) for details.

4. CONVERGENCE OF THE POSTERIOR PROCESS

In this section we study the convergence properties of the posterior process, $\{\mu_t\}$, for arbitrary (i.e., not necessarily optimal) policies.

The main results of this section may be described as follows. In Theorem 4.1, we show that the posterior process always converges (in the weak topology of measures), with probability one. However, the limiting probability, μ_∞, may or may not be concentrated on the true parameter. Having established the general convergence result, we proceed by simplifying the model a little (in particular, we assume that the distribution of shocks is known, and further that $k = 1$, so that we have a simple regression equation $y = \alpha + \beta x + \varepsilon$). Under this simplification, we are able to provide some characterization of the limiting distribution. In particular, we show that if, for some ω in Ω, $x_t(\omega)$ does not converge, then the limiting posterior distribution, for that ω in Ω, is concentrated on the "true" parameter value. Alternatively, if $x_t(\omega)$ does converge, to some $x(\omega)$ say, the posterior process converges to a limiting probability with support a subset of the set $\{(\alpha', \beta'), : \alpha' + \beta' x(\omega) = \alpha + \beta x(\omega)\}$, where α, β represent the "true" parameter values. (If the posterior process is a sequence of normal distributions, then this implies that the limiting distribution has a singular variance-covariance matrix.)

First we prove that under the very general conditions of Section 2 and 3 above,

the posterior process converges for P-a.e. ω in Ω, to a well-defined probability measure (with the convergence taking place in the weak topology).

Note that for any Borel subset, D, of the parameter space Θ, if we suppress the ω's and let, for some fixed ω, $\mu_t(D)$ represent the mass that measure $\mu_t(\omega)$ assigns to the set D, then

$$\mu_t(D) = E[1_{\{\theta \in D\}} | F_t] \tag{4.1}$$

Define a measure μ_∞ on Θ by setting, for each Borel set D in Θ,

$$\mu_\infty(D) = E[1_{\{\theta \in D\}} | F_\infty] \tag{4.2}$$

In the theorem below we indicate that μ_∞ is the limiting posterior distribution.

THEOREM 4.1. *The posterior process* $\{\mu_t\}$ *converges, for* P-a.e. ω *in* Ω, *in the weak topology, to the probability measure* μ_∞.

Interpreted in the context of our example, Theorem 4.1 establishes the convergence of the sequence of parameters of "beliefs," $\{\alpha_t, \beta_t, \Sigma_t\}$. Convergence in this specific context is established by Easley and Kiefer (1986) using direct calculations based on the updating formulas given in Section 2 above. Note that the theorem says nothing about what the sequence converges to: in particular the limiting means need not equal true values and the limit variance does not necessarily go to zero.

We now introduce a few simplifying assumptions, to enable us to characterize the convergence properties of the posterior process. These assumptions reduce the model to the situation of a simple regression equation. In particular, we suppose that Condition (*S*) holds:

CONDITION (S). *The shock process has a distribution which is known to the agent, and which possesses finite second moment; and* $k = 1$, *so that the action space* $\bar{X}$ *is a subset of* R^1.

In Theorem 4.2, we show that if the x_t process does not converge then the posterior process converges to the point mass on the true parameter value. Note however that nonconvergence of the x_t process is not necessary for convergence of μ_t to point mass.

Let $B = \{\omega : x_t(\omega) \text{ does not converge}\}$, and let 1_θ be the point mass at θ.

THEOREM 4.2. *For* μ_0-a.e. θ *in* Θ, *the posterior process,* $\mu_t(\omega)$, *converges to* 1_θ, *for* P_θ-a.e. ω *in* B.

Define on B^C, the set where $x_t(\omega)$ converges, $x(\omega) = \lim_{t\to\infty} x_t(\omega)$. In Theorem 4.3, it is shown that if the x_t process does converge, to $x(\omega)$, the posterior process converges to a limiting probability with support a subset of the set $\{(\alpha', \beta') : \alpha' + \beta' x(\omega) = \alpha + \beta\, x(\omega)\}$, where α, β represent the "true" parameter values.

THEOREM 4.3. *For* μ_0-a.e. $\theta = (\alpha, \beta)$ *in* Θ, *the posterior process* $\mu_t(\omega)$, *converges to a limiting distribution,* $\mu_\infty(\omega)$, *with support a subset of the set* $\{(\alpha', \beta') : \alpha' + \beta' x(\omega) = \alpha + \beta x(\omega)\}$, *for* P_θ-a.e. ω *in* B^C.

Theorem 4.2 interpreted in terms of our example shows that, if the $\{x_t\}$ sequence does not converge then $\{\alpha_t, \beta_t\}$ converges to the true values and $\{\Sigma_t\}$ converges to the zero matrix. Theorem 4.3 shows that, if the $\{x_t\}$ sequence converges, to x say, then the limiting distribution is singular with support on the line $\alpha' + \beta' x = \alpha + \beta x$ where α and β are true values. Essentially, when x is fixed, one learns the mean of y corresponding to that value of x.

5. OPTIMIZATION, LIMITING BELIEFS AND POLICIES

In this section the implications of optimizing behavior for the sequences $\{\mu_t\}$ of beliefs and $\{x_t\}$ of actions are considered. We will be particularly concerned with the limiting or "long-run" beliefs and policy.

Note first that convergence of beliefs was established, in Theorem 4.1, for an arbitrary $\{x_t\}$ sequence (i.e., without taking into account the underlying maximization problem). It therefore makes sense to ask what action (or actions) $\bar{x}$ corresponds to the limiting beliefs μ_∞.

Theorem 5.1 establishes that the limit action is the action which maximizes single period reward for limit beliefs.

We shall now impose the following strict concavity assumption on the reward function:

(A.2) For all priors μ in $P(\Theta)$, $r(x, \mu)$ is strictly concave in x.

THEOREM 5.1. *The limit action* $\bar{x} = \lim_{t\to\infty} x_t$ *exists, is unique (for given* μ_0*) and maximizes the one-period reward,* $r(x, \mu_\infty)$, *for limit beliefs* μ_∞.

Using Theorem 5.1 with Theorem 4.2 we observe that the optimal action converges (to $\bar{x}$) and the agent learns the value of $\alpha + \beta\bar{x}$. It is tempting to conclude that the agent has solved the information problem because even though the agent does not know α and β, the quantity $\alpha + \beta\bar{x}$ is known. The question however is whether $\bar{x}$ is indeed the correct action to take. The example of Section 6 below indicates that $\bar{x}$ may be the "wrong" action.

6. EXAMPLES

There are a handful of examples in the literature in which incomplete learning is optimal. The most familiar is no doubt the bandit framework used by Rothschild (1974) to show that a firm which could charge one of two prices each period, and faced an unknown purchase probability corresponding to each price, could end up charging the wrong price infinitely often. The classic bandit problem does not fit into our framework but the flavor of the result is similar. McLennan (1984) studied an example in which a monopolist knew the unknown demand curve he faced took

one of two values. In this example it can be optimal for the expected discounted profit maximizing monopolist to fail to learn which curve he faces. This example fits exactly into our framework, though it involves the special case of a discrete parameter space. Kihlstrom, Mirman and Postlewaite (1984) construct a related example involving a discrete parameter space.

Here we give a set of sufficient conditions in which incomplete learning is optimal in a problem with a continuous parameter space. The argument proceeds as follows: first, we specify a simple (quadratic) utility function, and using the normal data generating process and normal conjugate prior calculate the reward function. We then look for a candidate limiting belief-policy pair which does not exhibit complete learning. This involves finding an action $\bar{x}$ and a set of beliefs μ such that $\bar{x}$ maximizes $r(\cdot, \mu)$ and μ satisfies $\mu = \Gamma(\bar{x}, y, \mu)$ for all y. In addition μ must satisfy the constraint that the agent knows $\alpha^* + \beta^*\bar{x}$ where α^*, β^* are the true values. In our example there are many such belief-policy pairs. Finally, we verify under an additional assumption that our candidate belief-policy pair is optimal. The normal-normal model used in this example is studied in detail by Easley and Kiefer (1986).

Example of a Stationary One-Period Maximizing Policy. Consider the utility function

$$u(x, y) = -(y - y^*)^2 - x^2$$

where $y = \alpha + \beta x + \varepsilon$ and $\varepsilon \sim n(0, 1)$. Let $\mu = n(m, \Sigma)$ where $m = (\bar{\alpha}, \bar{\beta})'$ is the vector of means and $\Sigma = \{\sigma_{ij}\}$, $i, j = \alpha, \beta$ is the covariance matrix. The reward function is then

$$r(x, \mu) = -1 - y^{*2} - \sigma_{\alpha\alpha} - x^2\sigma_{\beta\beta} - 2x\sigma_{\alpha\beta} - \bar{\alpha}^2 - \bar{\beta}^2x^2 - 2\bar{\alpha}\bar{\beta}x + 2y^*\bar{\alpha} + 2y^*\bar{\beta}x - x^2$$

and the policy which maximizes current reward is $\bar{x} = (y^*\bar{\beta} - \bar{\alpha}\bar{\beta} - \sigma_{\alpha\beta})/(1 + \bar{\beta}^2 + \sigma_{\beta\beta})$. Looking back to the updating formulas in Section 2 we note that the requirement that $\mu = \Gamma(\bar{x}, y, \mu)$ for all y amounts to the constraint $(1\ \bar{x})\Sigma = 0$, or $\sigma_{\alpha\alpha} + \sigma_{\alpha\beta}\bar{x} = 0$ and $\sigma_{\alpha\beta} + \sigma_{\beta\beta}\bar{x} = 0$. Finally, the agent must learn the value of y corresponding to the limiting x, so our candidate solution $(\bar{x}, \mu)$ must satisfy $\bar{\alpha} + \bar{\beta}\bar{x} = \alpha^* + \beta^*\bar{x}$ where α^* and β^* are true values.

Given values of α^* and β^* we can find $(\bar{x}, \mu)$ pairs satisfying these conditions. One example is

(6.1a, b) $$\sigma_{\beta\beta} = 1, \qquad \sigma_{\alpha\beta} = -(y^* - \alpha^*)/(\beta^* + 1),$$

(6.1c, d) $$\sigma_{\alpha\alpha} = (y^* - \alpha^*)^2/(\beta^* + 1)^2, \qquad \bar{\beta} = 1,$$

(6.1e) $$\bar{\alpha} = \alpha^* + [(\beta^* - 1)(y^* - \alpha^*)/(\beta^* + 1)]$$

(6.1f) $$\bar{x} = (y^* - \alpha^*)/(\beta^* + 1).$$

We now demonstrate that running the policy in 6.1f is optimal for an agent whose beliefs are given by 6.1, for a sufficiently low discount factor under the assumption

that $EV(\Gamma(x, y, \mu))$ is twice differentiable with bounded (above) second derivative in x. This assumption is strong (though it can be relaxed somewhat) and should be shown from first principles, but we are currently unable to do so in general. The assumption can be shown to hold when the parameter space is discrete. Consider the function $\phi(x, \mu) = r(x, \mu) + \delta EV(\Gamma(x, y, \mu))$; here the distributions μ are indexed by the parameters m and Σ, thus derivatives with respect to μ can be worked through with the usual rules, but for the present purposes it suffices to proceed purely informally and write expressions like $V' = (dV/d\mu)$. Clearly,

$$V(\mu) = \max_{x \in \bar{X}} \phi(x, \mu).$$

The first order condition for the maximization problem is

$$\phi_x(x, \mu) = r_x(x, \mu) + \delta EV'(\Gamma(x, y, \mu))\Gamma_x = 0$$

where we have assumed that one can pass the derivative inside the expectation. Now, $\Gamma(x, y, \mu)$ evaluated at (x, μ) satisfying 6.1 does not depend on y, consequently V' can be taken outside the expectation. But new information is not expected to change current beliefs (if you expect your prior probabilities to change in a particular way you do not have the right prior), consequently $E\Gamma_x = 0$. (Proof: $E\Gamma_x = d/dx\, E\Gamma = d/dx\, \mu = 0$.) Solutions (x, μ) satisfying 6.1 also satisfy $r_x(x, \mu) = 0$ so the first-order conditions are satisfied for these solutions. The second derivative is

$$\phi_{xx}(x, \mu) = r_{xx}(x, \mu) + \delta \frac{d^2}{dx^2} EV(\Gamma(x, y, \mu)).$$

Now r_{xx} does not depend on current x (it is -4 for the values in 6.1). Consequently, as long as d^2EV/dx^2 is bounded above, δ can be chosen sufficiently small so that the function $\phi(x, \mu)$ is concave in x. Thus solutions to the first-order conditions (e.g., a solution satisfying 6.1) are indeed optimal. Note further that this action may be "wrong" in the sense that if the true values were known, the true values may be such that the agent may choose another action.

A variation on the technique used in this example shows how one can generate for a strictly concave utility function a prior such that for all sufficiently small discount factors it is optimal to choose the one-period optimal action which has no information value and may also be "wrong" in the sense explained above.

7. CONCLUSION

We have analyzed the problem faced by an agent attempting to control a linear regression process when the parameter values are unknown. In this setting the agent is faced with a tradeoff: by varying the values of the regressors he can accumulate information about the unknown coefficients, at a cost in terms of expected current utility. How much experimentation is appropriate? We restrict our attention to the problem facing a single decision maker, but the answer is

clearly of interest to the question of the amount of information generated by movements in economic variables when some agents have market power.

We show that the problem can be brought into the dynamic programming framework and that the value function satisfies the usual functional equation. We note that the optimal policies are typically different from the least-squares policies, and therefore offer an improvement over least-squares. We then turn our attention to the question of learning. To this end, we examine the asymptotic properties of the sequence of beliefs about the unknown parameters, that is, to the sequence of posterior distributions. This sequence is shown to converge almost surely. The limit distribution is not necessarily point mass at true values, nor indeed centered (in the sense of means) at the true values. In Section 4 we indicated that for an arbitrary (i.e., not necessarily optimal) action process there will be complete learning of the true parameter values if the action process does not converge, and there may be some (but possibly incomplete) learning if the action process does converge. In Section 5 we showed that the optimal action process converges (to the one-period optimal action under the limiting posterior distribution); hence we can not conclude anything about the learning of the true parameter vector in this case from the results of Section 4. From the example of Section 6, one can, in general, always find a prior probability on the parameter vector such that for sufficiently low discount factor the agent will choose the same totally uninformative action in each period. Hence at the level of generality considered in this paper, it seems that nothing more can be said about the question of learning.

APPENDIX

The Bayes' Rule Operator.

Let $P(dy_t\, d\theta | x_t, \mu_{t-1})$ be the joint distribution on $R^1 \times \Theta$ obtained as follows: an element θ in Θ is first chosen according to the probability μ_{t-1}; then, given this chosen value of $\theta = (\alpha, \beta, \phi)$, y_t is chosen according to the relation, $y_t = \alpha + \beta x_t + \varepsilon_t$, where ε_t has the distribution $p(\cdot | \phi)$. Next, define $P(dy_t | x_t, \mu_{t-1})$ to be the marginal distribution of $P(dy_t\, d\theta | x_t, \mu_{t-1})$ on R^1. We now apply Parthasarathy (1967, Ch. *V*, Theorem 8.1), to obtain the existence of a conditional probability measure on Θ, $\Gamma(d\theta | x_t, y_t, \mu_{t-1})$, which, for fixed (x_t, μ_{t-1}), is measurable in y_t, and where

$$P(dy_t\, d\theta | x_t, \mu_{t-1}) = P(dy_t | x_t, \mu_{t-1}) \cdot \Gamma(d\theta | x_t, y_t, \mu_{t-1}).$$

The conditional probability, $\Gamma(d\theta | x_t, y_t, \mu_{t-1})$, defines the Bayes' rule operator, $\Gamma(x_t, y_t, \mu_{t-1})$.

The Random Variables $\{x_t, y_t, \mu_t\}$.

We now provide the technical details behind the construction of the $\{x_t, y_t, \mu_t\}$ processes. Recall $(\Omega, F, P) = (\Theta, \bar{\Theta}, \mu_0) \times (\Omega', F', P')$. Any policy, π, generates a sequence of random variables $\{(x_t, y_t, \mu_t)\}_{t=1}^{\infty}$ on (Ω, F, P) as follows: first consider $\{\varepsilon_t\}$ as a stochastic process on (Ω, F, P), rather than (Ω', F', P'), by

$\varepsilon_t(\omega) = \varepsilon_t(\omega')$ where ω' is the second coordinate of ω (recall $\Omega = \Theta \times \Omega'$). μ_0 is given a priori; define $x_1(\omega) = \pi_0(\mu_0)$, $y_1(\omega) = \alpha + \beta x_1(\omega) + \varepsilon_1(\omega)$ and $\mu_1(\omega) = \Gamma(x_1(\omega), y_1(\omega), \mu_0)$, where α and β are obtained from the first coordinate of ω (recall $\Omega = \Theta \times \Omega'$). Since both π_0 and Γ are Borel functions (recall Γ is continuous), we observe that x_1, y_1 and μ_1 are (Borel measurable) random variables on (Ω, F, P).

Next, if we suppose that the random variables x_i, y_i and μ_i have been defined for $i = 1, \ldots, t-1$, then we may define, inductively, x_t, y_t, μ_t by putting $h_t(\omega) = (\mu_0; (x_1(\omega), y_1(\omega), \mu_1(\omega)), \ldots, (x_{t-1}(\omega), y_{t-1}(\omega), \mu_{t-1}(\omega)))$ and $x_t(\omega) = \pi_t(h_t(\omega))$, $y_t(\omega) = \alpha + \beta x_t(\omega) + \varepsilon_t(\omega)$ and $\mu_t(\omega) = \Gamma(x_t(\omega), y_t(\omega), \mu_{t-1}(\omega))$ Since π_t is Borel measurable, x_t, y_t, μ_t are (measurable) random variables.

Proofs.

PROOF OF THEOREM 4.1. Let U be the subclass of $\tilde{\Theta}$ made up of sets of the following kind: first, since Θ is separable, let $\{s_1, s_2, s_3, \ldots\}$ be a separant; let B_n^k be the ball of radius $1/n$ and center s_k; then define U to be the set of all finite intersections of the balls B_n^k, where $k = 1, 2, \ldots$ and $n = 1, 2, \ldots$. One may check that U is countable.

Next, for any fixed set D, $\mu_t(D) = E[1_{\{\theta \in D\}} \mid F_t]$, so using Chung (1974, Theorem 9.4.8, *p.* 340), the sequence $\{\mu_t(D)\}$ can be shown to be a positive Martingale, and so the Martingale convergence theorem applies and we conclude that $\mu_t(D)$ converges with P probability one to $\mu_\infty(D)$. Since the set U is countable, we have that convergence holds on all of U, simultaneously, with P probability one. Then we check that U satisfies conditions (i) and (ii) of Billingsley (1968, Theorem 2.2, p. 14), so, from that Theorem, μ_t converges weakly with P probability one.

Hence the limit of μ_t exists (a.e.). In the above paragraph we identified this limit with μ_∞ on all sets D in the class U. Since every open set in Θ is countable union of sets in U, and since open sets generate the σ-algebra $\tilde{\Theta}$, it must be the case that the limit of μ_t equals μ_∞ (on all sets in $\tilde{\Theta}$). Q.E.D.

Comment on Proof of Theorem 4.2. The idea behind the proof of Theorem 4.2 is the following. Suppose first that $x_t(\omega) = x'$ for all t and for all ω. Then $y_t(\omega) = \alpha + \beta x' + \varepsilon_t(\omega)$, and $\Sigma_{t=1}^n y_t(\omega)/n = \alpha + \beta x' + \Sigma_{t=1}^n \varepsilon_t/n$. However, by the strong law of large numbers, $\lim_{n\to\infty} \Sigma_{t=1}^n \varepsilon_t/n = 0$, P-a.e., so if we define $y' = \lim_{n\to\infty} \Sigma_{t=1}^n y_t(\omega)/n$, then $y' = \alpha + \beta x'$; and, in particular, the agent will learn that the true parameter will satisfy this relation, in the limit. Next, if $x_t(\omega)$ does not converge, but alternates between two numbers, x' and x'', it is obvious that applying the above argument first to the even sequence $\{x_{2t}\}_{t=1}^\infty$ and then to the odd sequence $\{x_{2t-1}\}_{t=1}^\infty$, the two equations $y' = \alpha + \beta x'$ and $y'' = \alpha + \beta x''$ will be obtained, where $y' = \lim_{n\to\infty} 1/n \Sigma_{t=1}^n y_{2t}$ and $y'' = \lim_{n\to\infty} 1/n \Sigma_{t=1}^n y_{2t-1}$ from which one may compute the true parameters, α and β. In this situation the agent will learn the true parameters in the limit. It is this idea which is behind the proof presented below.

In the example above, notice we had to apply the law of large numbers first to the even time subsequence and then to the odd subsequence. In Lemma 4.2 below, we show that the law of large numbers may be applied to a very large set of time

subsequences. The rest of the proof involves setting up the machinery to use Lemma 4.2.

PROOF OF THEOREMS 4.2 AND 4.3. Define $1_{\{\omega \in k\}} = 1$ if $\{\omega \in K\}$ and equal to zero otherwise, where K is any subset of Ω; we sometimes also write this as 1_k.

LEMMA 4.2. *There exists a set A in F with $P(A) = 1$, such that on for all rational numbers ℓ and m, with $\ell < m$, on the set where*

$$\Sigma_{t=1}^{\infty} 1_{\{\ell \leq x_t \leq m\}} = \infty,$$

$$\lim_{T \to \infty} \Sigma_{t=1}^{T} \varepsilon_t 1_t / \Sigma_{t=1}^{T} 1_t = 0 \tag{A.1}$$

where $1_t = 1_{\{\ell \leq x_t \leq m\}}$.

PROOF. Fix an ℓ and m with $\ell < m$. Note that for $t' \geq t$, ε_t, is independent of $\{1_1, \ldots, 1_t\}$ where $1_j = 1_{\{\ell \leq x_j \leq m\}}$. Hence, from Lemma 3.3 below, we obtain that on some set $A(\ell, m)$ with $P(A(\ell, m)) = 1$, (A.1) holds. Define A to be the intersection over all rational numbers $\ell < m$, of the sets $A(\ell, m)$; then $P(A) = 1$ and A satisfies the conclusion of the Lemma. Q.E.D.

LEMMA 4.3. *Let $\{v_t\}$ be a sequence of independent random variables with mean zero and uniformly bounded variance. Let $\{z_t\}$ be a sequence of random variables such that for each t, t' with $t' \geq t$, $v_{t'}$ is independent of $\{z_1, \ldots, z_t\}$; then for almost every realization with $\Sigma_{t=1}^{T} z_t^2 \to \infty$,*

$$\lim_{T \to \infty} \Sigma_{t=1}^{T} z_t v_t / \Sigma_{t=1}^{T} z_t^2 = 0. \tag{A.2}$$

PROOF. One applies Taylor (1974, Lemmas 1 through 3) with minor modifications.

PROOF OF THEOREMS 4.2 AND 4.3 (continued). To ease the exposition we shall assume that $\bar{X} = [0, 1]$; since $\bar{X}$ is assumed compact this is without loss of generality.

Let $\bar{Q}$ be the set of rational numbers in $\bar{X}$, and let $\bar{x}(\omega) = \limsup x_t(\omega)$ and $\underline{x}(\omega) = \liminf x_t(\omega)$. We proceed to define two random variables $h(\omega)$ and $h'(\omega)$ taking values in $\bar{Q}$ and such that on B, the set where x_t does not converge, $\underline{x}(\omega) < h'(\omega) < h(\omega) < \bar{x}(\omega)$. Define the function $h: \bar{X} \times \bar{X} \to \bar{Q}$, as follows. First, any integer $k = 1, 2, \ldots$, can be uniquely written as $k = 2^{n-1} + p$, where $n = 1, 2, \ldots$, and $0 \leq p \leq 2^{n-1} - 1$; so define $s_k = (p + 1)/2^{n-1}$, where $k = 2^{n-1} + p$. The sequence $\{s_k\}$ is therefore a sequence of rational numbers in $\bar{X} = [0, 1]$. Define $t(x, x') = \inf \{k: s_k \in (x, x')\}$ if $x < x'$, and $t(x, x') = 0$ if $x \geq x'$; and $h(x, x') = s_{t(x,x')}$ with $s_0 = 0$. Hence h takes values in $\bar{Q}$, and one can check that h is Borel measurable. (In fact, to prove the measurability of h, note that $t(x, x') = 1_{\{x < x'\}} \Sigma_{t=1}^{\infty} r_t$ where r_t is the indicator function which equals 1 when $s_k \in (x, x')^c$ for all integers $k < t$ and $s_t \in (x, x')$ and zero otherwise (with $r_1 = 1$). Since $1_{\{x > x'\}}$

and r_t (for each t) are Borel measurable, we obtain the measurability of $t(x, x')$; the measurability of h then follows from $h(x, x') = s_{t(x,x')} = \sum_{k=1}^{\infty} s_k 1_{\{t(x,x')=k\}}$.

Next, we define the random variable $h(\omega) = h(\underline{x}(\omega), \bar{x}(\omega))$ (note the abuse of notation !). Since $\bar{x}$ and $\underline{x}$ are both F_∞-measurable and $h(x, x')$ is Borel-Measurable, we obtain that $h(\omega)$ is F_∞-measurable. We have therefore constructed an F_∞-measurable random variable, $h(\omega)$, taking values in $\bar{Q}$, and such that on B, $\underline{x}(\omega) < h(\omega) < \bar{x}(\omega)$. By replacing $\bar{x}(\omega)$ with $h(\omega)$, and repeating the above construction, we obtain an F_∞-measurable random variable, $h'(\omega)$, taking values in $\bar{Q}$, and such that on B, $\underline{x}(\omega) < h'(\omega) < h(\omega)$.

The true parameter vector satisfies

$$y_t = \alpha + \beta x_t + \varepsilon_t \tag{A.3}$$

Define $\bar{1}_t = 1_{\{h \leq x_t \leq 1\}}$. Multiplying both sides of (A.3) by $\bar{1}_t$ summing over t and dividing by $\bar{S}_T = \sum_{t=1}^{T} \bar{1}_t$, one obtains

$$\bar{y}_T = \alpha + \beta \bar{x}_T + \bar{\varepsilon}_T \tag{A.4}$$

where $\bar{y}_T = \sum_{t=1}^{T} y_t \bar{1}_t / \bar{S}_T$, $\bar{x}_T = \sum_{t=1}^{T} x_t \bar{1}_t / \bar{S}_T$ and $\bar{\varepsilon}_T = \sum_{t=1}^{T} \varepsilon_t \bar{1}_t / \bar{S}_T$. From the definition of h note that $h \leq x_t \leq 1$ infinitely often, hence $\sum_{t=1}^{\infty} \bar{1}_t = \infty$; so from Lemma 4.2, $\bar{\varepsilon}_T \to 0$ as $T \to \infty$ a.e. on B. Taking the lim sup on both sides of (A.4) results in

$$\bar{y} = \alpha + \beta \bar{x}' \tag{A.5}$$

where $\bar{y} = \limsup \bar{y}_T$ and $\bar{x}' = \limsup \bar{x}_T$. Repeating the above exercise, but replacing $\bar{1}_t$ with $\underline{1}_t = 1_{\{0 \leq x_t \leq h'\}}$, we obtain that (a.e. on B)

$$\underline{y} = \alpha + \beta' \underline{x}' \tag{A.6}$$

where $\underline{y} = \limsup \underline{y}_T$, $\underline{y}_T = \sum_{t=1}^{T} y_t \underline{1}_T / \underline{S}_T$, $\underline{S}_T = \sum_{t=1}^{T} \underline{1}_t$ and $\underline{x}'$ is defined analagously. Hence if we define $M = \{(\alpha', \beta') : \bar{y} = \alpha' + \beta' \bar{x}' \text{ and } \underline{y} = \alpha' + \beta' \underline{x}'\}$ *since* $\underline{x}' < h' < h < \bar{x}'$, $\underline{x}' \neq \bar{x}'$ *so* M consists of only one point, which from (A.5) and (A.6) must be the true parameter vector $\theta = (\alpha, \beta)$ so $1_{\{\theta \in M\}} = 1$ (a.e. on B). Since clearly $1_{\{\theta \in M\}}$ is F_∞-measurable

$$\begin{aligned} \mu_\infty(M) \cdot 1_B &= E[1_{\{\theta \in M\}} | F_\infty] \cdot 1_B = E[1_{\{\theta \in M\}} \cdot 1_B | F_\infty] \\ &= E[1_B | F_\infty] = 1_B \text{ a.e.} \end{aligned} \tag{A.7}$$

From (A.7), therefore, on B the limiting posterior distribution is concentrated on the true parameter vector (a.e.). This proves Theorem 4.2.

To prove Theorem 4.3, we repeat the above exercise replacing $\bar{1}_t$ (or $\underline{1}_t$) with $1_t = 1_{\{0 \leq x_t \leq 1\}}$ (which of course is equal to one!). On B^c, $\lim_{t \to \infty} x_t = x$ (say) so if $M = \{(\alpha', \beta') : \bar{y} = \alpha' + \beta' x\}$, where $\bar{y} = \lim_{n \to \infty} 1/n \sum_{t=1}^{n} y_t$, then the true parameter vector $\theta = (\alpha, \beta)$ lies in M, and we may show, as in (A.7) that

$$\mu_\infty(M) \cdot 1_{B^c} = 1_{B^c} \text{ a.e.} \tag{A.7'}$$

from which Theorem 4.3 follows. Q.E.D.

PROOF OF THEOREM 5.1. Recall from Theorem 4.1 that $\lim_{t\to\infty} \mu_t = \mu_\infty$ exists for all sample paths. The sequence $\{x_t\}$ and $\{\mu_t\}$ satisfies for each t (simultaneously, a.e.) the functional equation

$$V(\mu_{t-1}) = r(x_t, \mu_{t-1}) + \delta \cdot \int V(\Gamma(x_t, y_t, \mu_{t-1}))p(d\varepsilon \mid \phi)\mu_{t-1}(d\alpha\ d\beta\ d\phi). \tag{A.8}$$

where $y_t = \alpha + \beta x_t + \varepsilon_t$. Taking limits along any convergent subsequence gives

$$V(\mu_\infty) = r(\bar{x}, \mu_\infty) + \delta \int V(\Gamma(\bar{x}, y, \mu_\infty))p(d\varepsilon \mid \phi)\mu_\infty(d\alpha\ d\beta\ d\phi) \tag{A.9}$$

where $\bar{x}$ is a limit point of the $\{x_t\}$ sequence. (In taking the limits one uses the fact that V is bounded and the integral in (A.8) is $E[V(\mu_t) \mid F_{t-1}]$ to apply Chung 1974, Theorem 9.4.8). However, from Theorems 4.2 and 4.3, if $\bar{y} = \alpha + \beta \bar{x} + \varepsilon$, the value $y'' = \alpha + \beta \bar{x}$ is known (under F_∞), hence $\bar{y} = \alpha + \beta \bar{x} + \varepsilon = y'' + \varepsilon$ becomes a white noise term (with mean y''); hence observing $(\bar{x}, \bar{y})$ yields no information so $\Gamma(x, y, \mu_\infty) = \mu_\infty$, and (A.9) becomes

$$V(\mu_\infty) = r(\bar{x}, \mu_\infty) + \delta V(\mu_\infty). \tag{A.10}$$

Now we show that $\bar{x}$ solves the problem

$$\max_{x \in \bar{X}} r(x, \mu_\infty) \tag{A.11}$$

Suppose on the contrary that there is an $\hat{x} \in \bar{X}$ such that

$$r(\bar{x}, \mu_\infty) > r(\bar{x}, \mu_\infty). \tag{A.12}$$

Using the functional equation

$$V(\mu_\infty) \geq r(\hat{x}, \mu_\infty) + \delta \int V(\Gamma(\hat{x}, \hat{y}, \mu_\infty))P(d\varepsilon \mid \phi)\mu_\infty(d\alpha\ d\beta\ d\phi). \tag{A.13}$$

From Blackwell's Theorem (see e.g., Kihlstrom (1974, Lemma 1, p. 18)), since the experiment "observe $(\hat{x}, \hat{y})$" is trivially sufficient for the experiment "make no observations," we obtain,

$$\int V(\Gamma(\hat{x}, \hat{y}, \mu_\infty))p(d\varepsilon \mid \phi)\mu_\infty(d\alpha\ d\beta\ d\phi) \geqq V(\mu_\infty). \tag{A.14}$$

Hence, from (A.12) through (A.14).

$$V(\mu_\infty) > r(\bar{x}, \mu_\infty) + \delta V(\mu_\infty), \tag{A.15}$$

which is a contradiction to (A.10). So $\bar{x}$ solves problem (A.11); that is, $\bar{x}$ maximizes the one-period reward $r(x, \mu)$ for limit beliefs, μ_∞. Since $r(\cdot, \mu_\infty)$ is a strictly concave in x, $\bar{x}$ must be unique. Q.E.D.

REFERENCES

ANDERSON, T. W. AND J. TAYLOR, "Some Experimental Results on and Statistical Properties of Least Squares Estimates in Control Problems," *Econometrica* 44 (1976), 1289–1302.

BILLINGSLEY, P., *Convergence of Probability Measures* (New York: Wiley, 1968).

BLACKWELL, D., "Discounted Dynamic Programming," *Annals of Mathematical Statistics* 36 (1965), 2226–2235.

BLUME, L. AND D. EASLEY, "Rational Expectations Equilibrium: An Alternative Approach," *Journal of Economic Theory* 34 (1984), 116–129.

CHOW, G. C., *Econometric Analysis by Control Methods* (New York: Wiley, 1981).

CHUNG, K. L., *A Course in Probability Theory*, 2nd edition (New York: Academic Press, 1974).

EASLEY, D. AND N. M. KIEFER, "Controlling a Stochastic Process with Unknown Parameters: Handout," manuscript, Cornell University, 1985.

——— AND ———, "Infinite Horizon Bayesian Control of the Normal-Normal Regression Process," working paper, Cornell University, 1986.

HARKEMA, R., "An Analytical Comparison of Certainty Equivalence and Sequential Updating," *Journal of the American Statistical Association* 70 (1975), 348–350.

KIHLSTROM, R. E., "A 'Bayesian' Exposition of Blackwell's Theorem on the Comparison of Experiments," M. Boyer and R. E. Kihlstrom, eds., *Bayesian Models in Economic Theory* (Amsterdam: Elsevier Science Publishers B.V., 1984).

———, L. J. MIRMAN AND A. POSTLEWAITE, "Experimental Consumption and the 'Rothschild Effect,' M. Boyer and R. E. Kilhstrom, eds., *Bayesian Models in Economic Theory* (Amsterdam: Elsevier Science Publishers B.V., 1984).

JORDAN, J. S., "The Strong Consistency of the Least Squares Control Rule and Parameter Estimates," manuscript, Cornell University, 1985.

MAITRA, A., "Discounted Dynamic Programming in Compact Metric Spaces," *Sankhya* 30 (Series A) (1968), 211–216.

MCLENNAN, A., "Price Dispersion and Incomplete Learning in the Long Run," *Journal of Economic Dynamics and Control* 7 (1984), 331–347.

PARTHASARATHY, K., *Probability Measures on Metric Spaces* (New York: Academic Press, 1967).

PRESCOTT, E., "The Multiperiod Control Problem under Uncertainty," *Econometrica* 40 (1972), 1043–1058.

ROTHSCHILD, M., "A Two-Armed Bandit Theory of Market Pricing," *Journal of Economic Theory* 9 (1974), 185–202.

SCHWARTZ, L., "On Bayes Procedures," *Z. Wahrscheinlichkeits-theorie* 4 (1965), 10–26.

TAYLOR, J. B., "Asymptotic Properties of Multiperiod Control Rules in the Linear Regression Model," *International Economic Review* 15 (1974), 472–484.

ZELLNER, A., *An Introduction to Bayesian Inference in Econometrics* (New York: Wiley, 1981).