Journal of Economic Dynamics and Control 15 (1991) 687–713. North-Holland

On the convergence of Bayesian posterior processes in linear economic models

Counting equations and unknowns

Yaw Nyarko*
New York University, New York, NY 10003, USA

Received June 1989, final version received October 1990

I propose a technique, counting 'equations' and 'unknowns', for determining when the posterior distributions of the parameters of a linear regression process converge to their true values. This is applied to examples and to the infinite-horizon optimal control of this linear regression process with learning, and in particular to the problem of a monopolist seeking to maximize profits with unknown demand curve. Such a monopolist has a tradeoff between choosing an action to maximize the current-period reward and to maximize the information value of that action. I use the above technique to determine the monopolist's limiting behavior and to determine whether in the limit it learns the true parameter values of the demand curve.

1. Introduction

Consider the problem of estimating the parameter vector $\Theta = \{\alpha, \beta_1, \dots, \beta_k\}$ in the regression equation $y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + \varepsilon_t$, where x_{it} is the value the ith regressor takes at date t and $\{\varepsilon_t\}$ is an unobserved shock process.

Suppose an observer of this process starts with a prior, μ_0, over the parameter $\Theta = \{\alpha, \beta_1, \dots, \beta_k\}$, obtains sequentially the samples $\{y_t, x_{1t}, x_{2t}, \dots, x_{kt}\}$ for $t \geq 1$, and updates the prior using Bayes' rule. The observer of this process may be an econometrician seeking to estimate the values of the parameter vector Θ. Alternatively, the observer may be an

*I thank Professors N. Kiefer and D. Easley, members of the N.Y.U. Theory Workshop, and the anonymous referees for very helpful comments. I also thank the C.V. Starr Center for its generous support.

agent in an economy (a monopolist say), in which case some of the regressors may be actions of the agent that must be chosen as a function of the posterior distribution at the beginning of the period to maximize some objective function over some time horizon.

When the observer of the regression process is an agent in the economy that also wants to control the regression process, then there is a tradeoff between maximizing current-period rewards and the future information value of any action. This problem has been studied by Prescott (1972) and Grossman et al. (1977) among the earlier papers in the economics literature and more recently by Kihlstrom et al. (1984), McLennan (1984, 1986), Aghion et al. (1986), Easley and Kiefer (1988), and Kiefer and Nyarko (1989).

An important question in these models is whether in the limit there will be complete learning of the true parameter vector $\Theta = \{\alpha, \beta_1, \ldots, \beta_k\}$ of the regression process. One of the first papers to study this (albeit in a slightly different context) is Rothschild (1974), who showed that an agent solving an infinite-horizon control problem may optimally decide not to take actions that will lead to complete learning of the true parameter vector. In the context of the linear regressions model, examples of incomplete learning have been obtained in the papers mentioned earlier. The question of learning is also related to work on the convergence to rational expectations equilibria [see, e.g., Blume and Easley (1984) and Feldman (1987)].

This paper focuses primarily on studying when the observer of the linear regression process will over time learn the true parameter vector. Our contribution is two-fold. First, we provide a simple technique for determining when there will be complete learning of the true parameter vector by counting the number of 'equations' and 'unknowns'. We also study the infinite-horizon optimal control of the linear regression process and relate the question of complete learning to the properties of the one-period (myopic) problem.

By the number of unknowns we mean the number of parameters to be estimated, $(k+1)$ in our case (i.e., the parameters $\alpha, \beta_1, \ldots, \beta_k$). Let X_t be the vector $(x_{1t}, \ldots, x_{kt})$. Let X' be any finite limit point of the vector X_t [e.g., when X_t is a scalar ($k=1$), then if X_t converges it will have only one limit point, while if X_t does not converge it will have at least two limit points, the liminf and the limsup). Given any finite limit point X', we show that the observer of the regression process will learn the value of $\alpha + \beta_1 x'_1 + \cdots + \beta_k x'_k$ ($= y'$, say) where $X' = (x'_1, \ldots, x'_k)$; i.e., the observer of the process learns one 'equation' that the true parameter vector will satisfy. Hence if the X_t process has $(k+1)$ finite limit points $X^1, \ldots, X^{k+1}$, such that the corresponding vectors $(1, X^1), \ldots, (1, X^{k+1})$ [where $(1, X^j) = (1, x^j_1, \ldots, x^j_k) \in R^{k+1}$ for each $j = 1, \ldots, k+1$] are linearly independent, then the agent will learn $k+1$ linearly independent equations that the $k+1$

unknowns satisfy. The agent therefore will learn over time the true parameter vector.

In section 4, we study the problem of the optimal control of a linear regression process. Here one of the regressors is a control variable of an economic agent. The control variable will in general depend upon all past observations and so will not be an independent process. We therefore cannot apply classical techniques to study the convergence to the true parameter. We use instead the method of counting equations and unknowns. The results of this section may be summarized as follows: Consider an agent acting myopically; that is, given the posterior distribution at any date, this agent chooses at that date an action to maximize the one-period reward. This generates a sequence of values of the vector of regressors $X_t = (x_{1t}, \ldots, x_{kt})$ some of which are the control variables. Suppose that this necessarily results in X_t having at least $k+1$ limit points $X^1, \ldots, X^{k+1}$ such that the corresponding vectors $(1, X^1), \ldots, (1, X^{k+1})$ are linearly independent. Then the myopic agent learns the true parameter vector; but this then implies that the agent that is acting optimally (taking the future into account) will necessarily learn the true parameter values.

The usefulness of this result of course lies in the fact that for most models the one-period (myopic) problem can be solved and the number of limit points of the regressors can very easily be determined. In section 4 we relate this to whether the agent observes an exogenous process before or after choosing an action for the period and whether the one-period optimal action is linear in the observation of the exogenous process. We then illustrate this using the example of a monopolist with unknown demand curve maximizing a discounted sum of profits over an infinite horizon.

This result is in some sense the best that can be obtained with the level of generality considered in this paper. For in Kiefer and Nyarko (1989) it has been shown that for sufficiently low discount factors (i.e., if the future is sufficiently unimportant) a risk-averse agent solving an infinite-horizon problem may actually choose the same actions as an agent acting myopically.

We end the paper in section 5 with some concluding remarks. Most proofs are relegated to the appendix.

Relationship to existing literature

The question of learning has also been studied in 'non-Bayesian' contexts by Frydman (1982), Bray and Savin (1986), and Marcet and Sargent (1989) among others. The models studied in those papers differ from that in this paper in two essential ways: first, the forecasts of the unknown parameter are made by using least squares point estimators; and second, the unknown parameters are not fixed over time (as in this paper) but may vary as a function of the agents' estimates of the parameter. The most general results in that literature are in Marcet and Sargent (1989) which proves the convergence of the least squares estimates to the rational expectations values under assumptions which require some key matrices in their linear data-generating process to have eigenvalues less than one in absolute value and which also require the agent to ignore data which sends the least-squares estimates outside some specified compact set. Using least squares estimates of course means agents are using a biased and incorrect statistical model whenever the economy is not in a rational expectations equilibrium.

This paper on the other hand studies a model where the unknown parameter is fixed over time but where the agents use the correct statistical model in learning. Hence the statements and proofs use probabilistic arguments which do not require any strong assumptions.

This paper borrows from Kiefer and Nyarko (1989). There a similar problem to that studied in section 4 below. There are two main types of results in that paper. The first is that whether learning occurs or not is a function of whether the single regressor in the model converges or not; and the second showed that under regularity conditions for all discount factors sufficiently small the agent may optimally choose at each date the myopic action, so may not learn the true parameter. The model studied there had only one regressor, and the proofs were direct proofs that stressed the particular structure of that model. In this paper, we emphasize the method of counting equations and unknowns. This allows us in the context of the optimal control problem of section 4 to extend greatly the types of models that can be analyzed to those which have more than one regressor and which also have other exogenous regressors (i.e., which are not choice variables of the agent.) Further, the technique of counting equations and unknowns may be of some use in large-sample Bayesian statistics, as illustrated in the examples of section 3.

We stress that the results of this paper do not require any assumptions on the distributional form of the shock process or the prior distribution (e.g., normality assumptions), as is sometimes required in the econometrics or large-sample Bayesian statistics literature. In particular we do not require that the shock process or the prior be normally distributed or that the prior belong to a conjugate family.

One method for obtaining convergence of Bayesian posterior process to the true parameter value is to obtain conditions under which some estimator [e.g., ordinary least squares (O.L.S.)] is consistent and then to use a result which says that if there exists a consistent estimator, then the Bayesian posterior process converges. However, the conditions required to check the consistency of the O.L.S. estimator are in many situations harder or less general than merely counting the number of 'equations and unknowns' along any given sample path (and we illustrate this at the end of section 2 below.)

The results of section 4 are related to Easley and Kiefer (1988) who study the general dynamic programming problem with learning and relate the limiting optimal actions to the one-period optimal actions. In contrast, in section 4 we study models with more structure and where learning is through a linear regression equation to obtain, via the method of counting equations and unknowns more precise conclusions for these cases.

2. The number of equations and unknowns

Suppose a regression equation is given by

$$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + \varepsilon_t, \tag{1}$$

where x_{it} in R^1 is the value the ith regressor takes at date t and $\{\varepsilon_t\}$ is an unobserved independent sequence with known[1] distribution which has zero mean and uniformly bounded second moments. In particular, ε_t is independent of $\{(\varepsilon_1, x_1), \ldots, (\varepsilon_{t-1}, x_{t-1}), x_t\}$ while x_t can depend upon $\{(\varepsilon_1, x_1), \ldots, (\varepsilon_{t-1}, x_{t-1})\}$. Let μ_0 be the prior probability over the parameter vector $\Theta = \{\alpha, \beta_1, \ldots, \beta_k\}$ in R^{k+1} representing initial beliefs about Θ; we assume that μ_0 has finite first moment.

Let (Ω, F, P) be the probability space on which we define the random variables $\{x_{it}, \varepsilon_t, y_t\}$ for $i = 1, \ldots, k$ and $t \geq 1$. Denote by μ_t the posterior distribution at the end of date t, that is, after observing the samples $\{y_\tau, x_{1\tau}, \ldots, x_{k\tau}\}_{\tau=1}^{t}$. We can show that with probability one, the posterior distribution, μ_t, converges to some limiting distribution, μ_∞. In particular, we show:

Lemma 2.1. On almost every sample path, the posterior process, $\{\mu_t\}$, converges in the weak topology of measures to a probability measure, μ_∞.
(All proofs are in the appendix.)

It must be stressed that the limiting value of the posterior process, μ_∞, will depend on the sample path. Thus μ_∞ should be regarded as a random function: a mapping from Ω, the set of sample paths, to the set of probability distributions for Θ.

Lemma 2.1 indicates that the posterior process converges, and therefore settles down somewhere. The lemma however does not tell us where the process converges to. In particular the lemma does not say that the posterior

[1] We may relax the assumption that the error term has known distribution and assume only that the likelihood function used in updating assigns full mass to ε_t having zero mean. This is because the proof of the main theorem uses the strong law of large numbers. However, if the distribution of ε is unknown, then care must be taken in defining the correct probability space and attention must be paid to what happens on probability zero sets.

process converges to the probability measure concentrated on the true parameter value, Θ. We proceed to describe a method by which one can characterize the limiting posterior distribution, μ_∞, by counting the 'number of equations' and the 'number of unknowns', where the number of unknowns is $k+1$ (i.e., the parameters $\alpha, \beta_1, \ldots, \beta_k$) and the number of equations is the number of linearly independent limit points of the vector $(1, X_t) = (1, x_{t1}, \ldots, x_{tk})$.

Fix any sample path and consider the sequence $\{X_t\}_{t=1}^\infty$ where $X_t = (x_{1t}, \ldots, x_{kt})$; for that fixed sample path the sequence $\{X_t\}_{t=1}^\infty$ is a sequence of vectors in R^k. If there exists a sub-sequence of dates $\{t_n\}_{n=1}^\infty$ such that as $n \to \infty$ the limit of the vector $X_{t_n} = (x_{1t_n}, \ldots, x_{kt_n})$ exists and equals some $X' = (x'_1, \ldots, x'_k)$, then we say the sequence $\{X_t\}_{t=1}^\infty$ has X' as a limit point. We show in Theorem 2.2 below that if the sequence has X' as a limit point along the fixed sample path, there will exist a number y' known to the agent in the limit (along the same fixed sample path), such that the agent will learn that the true parameter vector $\Theta = (\alpha, \beta_1, \ldots, \beta_k)$ satisfies the relation $y' = \alpha + \beta_1 x'_1 + \cdots + \beta_k x'_k$. This results in the agent learning one 'equation' that the true parameter satisfies.

Before we state Theorem 2.2, we require some notation. Recall that μ_0 denotes the prior distribution over the true parameter $\Theta = (\alpha, \beta_1, \ldots, \beta_k)$. The parameter Θ may be considered a random variable chosen at 'date 0' according to the distribution μ_0.

Let F_t denote the σ-field generated by the random variables $\{y_\tau\}_{\tau=1}^{t-1}$ and $\{x_{1\tau}, \ldots, x_{k\tau}\}_{\tau=1}^{t}$. Note that for each t, $X_t = (x_{1t}, \ldots, x_{kt})$ is F_t-measurable but y_t is not. The σ-algebra F_t represents the information available to the observer of the regression process (1) at date t, just before the random variable y_t is observed. Finally let $F_\infty = \vee_{n=1}^\infty F_n$, the σ-algebra induced by the union of the F_n's, $\cup_{n=1}^\infty F_n$; F_∞ represents the information available to the agent in the limit.

We now state Theorem 2.2.

Theorem 2.2. There exists a set A of sample paths with $P(A) = 1$ with the following property: For each fixed sample path in A, if there exists a sub-sequence of dates $\{t_n\}_{n=1}^\infty$ such that as $n \to \infty$ the limit of the vector $X_{t_n} = (x_{1t_n}, \ldots, x_{kt_n})$ exists and equals $X' = (x'_1, \ldots, x'_k)$, then there exists an F_∞-random variable[2] y' such that the limiting posterior distribution, μ_∞, has support on the set

$$M = \{(\alpha', \beta'_1, \ldots, \beta'_k) \colon y' = \alpha' + \beta'_1 x'_1 + \cdots + \beta'_k x'_k\}.$$

Further the true parameter vector $\Theta = (\alpha, \beta_1, \ldots, \beta_k)$ lies in M.

[2]For technical reasons we require that the limit point, $X' = (x'_1, \ldots, x'_k)$, can be chosen to be a measurable function of the sample path.

Fix any sample path and consider the sequence $\{X_t\}_{t=1}^{\infty}$ where $X_t = (x_{1t}, \ldots, x_{kt})$; these are vectors in R^k. Such a sequence of vectors can have any number of limit points (and we stress that this is along the same given fixed sample path).

Suppose that for the *fixed* sample path the sequence of vectors $\{X_t\}_{t=1}^{\infty}$ has $k+1$ limit points $X^{(1)}, \ldots, X^{(k+1)}$, where for each $j = 1, \ldots, k+1$, $X^{(j)} = (x_1^j, \ldots, x_k^j)$. Then the agent, along that fixed sample path, will learn that the true parameter will satisfy $k+1$ equations, $y^j = \alpha + \beta_1 x_1^j + \cdots + \beta_k x_k^j$, say, where j varies from 1 to $k+1$. However there are $k+1$ 'unknowns', $\alpha, \beta_1, \ldots, \beta_k$. Hence if the $k+1$ equations that the agent learns are linearly independent, then there will only be one value of the parameter vector $\Theta = (\alpha, \beta_1, \ldots, \beta_k)$ that can satisfy all $k+1$ equations simultaneously. Since the true parameter value satisfies this relation, the agent will learn the true parameter in the limit.

(We emphasize that all of the above computations are being done along one given sample path. The agent is not expected to be able to make observations across different sample paths. However, since along a *given* sample path there may exist many limit points the agent may learn many 'equations' along the given sample path.)

The conditions for the convergence of the posterior process to the true value found in the econometrics and large-sample Bayesian statistics literature usually involve joint restrictions on both the distributional form of the priors and errors (e.g., normality). The method of counting equations and unknowns does not require such conditions.

One technique used in the literature in obtaining convergence of posterior distributions is to impose conditions under which the ordinary least squares (O.L.S.) estimator is consistent, and then to argue that if there exists a consistent estimator, then the Bayesian posterior process converges.

For example, if X_n denotes the matrix of regressors at date n, then an assumption typically imposed for strong consistency of the O.L.S. estimators is that $X_n' X_n / n$ converges to a finite positive definite matrix as $n \to \infty$ [see, e.g., White (1984, theorem 2.12)]. There are many cases where this condition does not hold and yet one can still apply the number of equations and unknowns technique (and we illustrate one such case in a footnote).[3] This condition may of course be weakened [see, e.g., Anderson and Taylor (1979)], but then the other conditions required for the case with multiple regressors

[3]Consider a linear regression model $y = \alpha + \beta x + \varepsilon$ with one regressor which takes the value $x_n = 1$ at any date $n = 2^k$ for any integer k, and $x_n = 0$ for all other dates. In this case Theorem 2.2 implies that on each sample path the posterior distribution converges to the true parameter vector (α, β). However, $\sum_{t=1}^n x_n^2/n$ converges to zero.

To see this, define for each integer n, $m(n)$ to be the unique integer such that $2^{m(n)-1} < n \leq 2^{m(n)}$. Then $\sum_{t=1}^n x_n^2/n = \sum_{t=1}^n x_n/n \leq \sum_{t=1}^{2^{m(n)}} x_n/2^{m(n)-1} = 2\sum_{t=1}^{2^{m(n)}} x_n/2^{m(n)} = 2m(n)/2^{m(n)} \to 0$ as $n \to \infty$.

are not always easy to directly verify and usually are much stronger than conditions required to apply the number of equations and unknowns technique.

Most important of all, of course, is the fact that the method of counting equations and unknowns yields results that hold on a *given* sample path; for example in the linear regression model with one regressor we conclude from Theorem 2.2 that the Bayesian posterior distribution converges on any given sample path that has at least two limit points.

3. Examples

We now provide a number of examples that show how Theorem 2.2 and the technique of counting the equations and unknowns may be used. Example 1 has i.i.d. regressors, while example 2 has a lagged dependent variable. We show that in both cases there is sufficient variability in the regressors (in particular, there are enough 'equations') and therefore there is complete learning of the true parameter vector. We stress that we place no assumptions on the particular distributional form of the prior distribution or the distribution of the shocks, $\{\varepsilon_t\}$; in particular we do not use the normal distribution or conjugate priors.

We later provide some counter-examples to show what may cause there to be incomplete learning in the long run (i.e., μ_∞ not concentrated on the true parameter vector). These are taken from the economics literature and involve cases where there is insufficient movement in the regressors (i.e., not enough 'equations').

Example 1 (the i.i.d. case). Suppose the following two conditions hold:

(i) The vector of regressors, $X_t = (x_{1t}, \dots, x_{kt})$, is serially independent and identically distributed.

(ii) For each t (recall that X_t is identically distributed) the random vector $(1, X_t) = (1, x_{1t}, x_{2t}, \dots, x_{kt})$ has at least $k+1$ linearly independent points in its support.

Condition (ii) holds if, for example, the random variable x_{it} is independent of x_{jt} for all i different from j and, in addition, for each i, x_{it} has at least two points in its support (i.e., x_{it} is not degenerate); in particular this will hold if $k = 1$ so that $y_t = \alpha + \beta x_t + \varepsilon_t$ and x_t has at least two points in its support. Condition (ii) also holds if X_t has R^k as its support (e.g., if X_t is multivariate normal). Condition (ii) fails when, for example, $x_{it} = x_{jt}$ for some i different from j; in this case there may be no way of learning β_i and β_j even though the sum, $(\beta_i + \beta_j)$, may be learnt.

Under the two conditions above we can show:

Proposition 3.1. Under (i) and (ii) above, outside of a null set there will be complete learning of the true parameter. [I.e., on almost every sample path the limiting posterior distribution, μ_∞, is concentrated on the true parameter vector, $\Theta = (\alpha, \beta_1, \ldots, \beta_k)$.]

The idea behind the proof of Proposition 3.1 is as follows: First, the assumption that X_t is serially independent and independent distributed implies, loosely speaking, that on each sample path the X_t process has as many limit points as there are points in the support of that process. In particular,

Lemma 3.2. Let $\{X_t\}$ be a serially independent and identically distributed process taking values in R^k. Let B be any Borel subset of R^k such that $P(X_1 \in B) > 0$. Then $P(X_t \in B$ infinitely often$) = 1$, and therefore outside of a null set X_t has a limit point in B.

Proof. Since the $\{X_t\}$ process is identically distributed,

$$\sum_{t=1}^{\infty} P(X_t \in B) = \sum_{t=1}^{\infty} P(X_1 \in B) = \infty. \tag{2}$$

The lemma then follows immediately from an application of the Borel–Cantelli lemma [see for example Chung (1974, p. 76)]. ■

To complete the proof of Proposition 3.1 one uses conditions (i) and (ii) and Lemma 3.2 to show that on almost every sample path the X_t process will have at least $(k+1)$ limit points, $X^1, X^2, \ldots, X^{k+1}$, say, [where for each j, $X^j = (x_1^j, \ldots, x_k^j)$ is a vector in R^k] such that the corresponding $k+1$ vectors $\{(1, X^1), \ldots, (1, X^{k+1})\}$ are linearly independent. One then uses Theorem 2.2 to conclude that the agent will learn $k+1$ linearly independent equations that the true parameter must satisfy, and hence will necessarily learn the true parameter vector.

Example 2 (lagged dependent variable). We now discuss the case where the regression equation is given by

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t, \tag{3}$$

where we assume that y_0 is a known and fixed, but otherwise arbitrary number.

We assume that the $\{\varepsilon_t\}$ process has mean zero and is not degenerate (i.e., is not concentrated on the point $\{0\}$). This is sufficient to induce enough variability in the y_t process so that on each sample path the $\{y_{t-1}\}$ process has at least two finite limit points, y' and y'', say. Using Theorem 2.2, we may conclude that there will exist some random variables, $\hat{y}'$ and $\hat{y}''$, known to the agent in the limit, such that the agent will learn that the true parameter vector (α, β) lies in both of the sets

$$M' = \{(\alpha', \beta') : \hat{y}' = \alpha' + \beta' y'\}, \tag{4}$$

$$M'' = \{(\alpha', \beta') : \hat{y}'' = \alpha' + \beta' y''\}. \tag{5}$$

Since there is only one value of α and β that can lie in M' and M'' simultaneously, the agent necessarily learns the true value of the parameters, α and β. For technical reasons we also require that the parameter β be of absolute value strictly less than unity, i.e., $|\beta| < 1$. Thus we can show:

Proposition 3.3. Suppose that the regression process is given by (3) above with $|\beta| < 1$. Then on almost every sample path the limiting posterior probability, μ_∞, will be concentrated on the true parameter vector, (α, β).

Counter-examples. The examples in the economics literature (mentioned in the introduction) showing that there may be incomplete learning of the true parameter vector all involve situations where there is insufficient variability in the regressors. For example, the model of Rothschild (1974) involves essentially the situation where an agent has to choose one of two arms of a bandit, both with unknown distribution of returns, so as to maximize an expected sum of discounted returns. Rothschild (1974) shows that it is possible that an agent after some finite sequence of draws will incorrectly predict that one of the arms is the 'good' arm and play this arm for all but infinitely many periods. Without variability in the use of the arms it is impossible to learn the true distribution of both arms.

4. An infinite-horizon optimal control model

The purpose of this section is to study what happens to the results of section 3, when instead of assuming the stochastic properties of the regressors we suppose that some of the regressors are obtained via an infinite-horizon optimization problem.

The second purpose of this section is to discuss to what extent the results obtained in Kiefer and Nyarko (1989) for the simple regression model (which showed that for all sufficiently small discount factors incomplete learning may be optimal) may be extended to the model where there are many regressors. By the simple regression model we mean the case where there is only one regressor as below (where x_t is the agent's control variable):

$$y_t = \alpha + \beta x_t + \varepsilon_t. \tag{6}$$

We now extend the model by introducing an exogenous regressor $\{z_t\}$ that we shall assume is serially independent and identically distributed. The regression equation then becomes

$$y_t = \alpha + \beta x_t + \phi z_t + \varepsilon_t. \tag{7}$$

We consider two cases which will depend upon whether the agent chooses an action at any date before (model A) or after (model B) the agent observes the value of the exogenous process for that date.

A priori it is difficult to say whether there will be more or less learning in these extended models. On the one hand, in the extended models there will be more parameters to learn (three parameters instead of two). However, recall that in the simple regression case the reason why there may be incomplete learning of the true parameter is because there is insufficient variability in the regressors. The extended models introduce some variability; the question is whether this added variability will be enough to estimate the larger number of parameters of the model.

In the long run, we know from Lemma 2.1 that the posterior process converges, and therefore settles down. In model A, where the agent chooses the action *before* observing the value of the $\{z_t\}$ process for the period, there will be no reason to expect the agent to move the optimal action sequence around if the posterior process is settling down. Hence, in the limit we expect the x_t process to settle down. In this situation it is not surprising that the agent may not learn the true parameter vector (especially the coefficient, β, of x in the regression equation). Proposition 4.2 discusses further the question of learning in model A.

In model B where the agent chooses a date t action, x_t, *after* the agent has observed the realization of z_t, since the posterior process is settling down, the optimal action x_t will depend very much on the observed value of z_t. The variations in the z_t process over time (recall z_t is i.i.d.) will therefore result in variations in the x_t process. The question of the complete learning of the true parameter vector will then depend upon whether in the limit the x_t process varies linearly or nonlinearly with the z_t process. For example if in the limit x_t is approximately equal to z_t, then the agent may learn $(\beta + \phi)$, but will not learn either β or ϕ.

In section 4.2, we identify the limits of the x_t process as the solution of the one-period problem with prior distribution equal to the limiting beliefs. The question of whether in the limit x_t moves linearly with z_t then depends upon whether or not the one-period optimal actions are linear or nonlinear in z

for fixed prior. Thus in model B there will be complete learning of the true parameter if the one-period problem results in optimal actions that are nonlinear functions of z; otherwise one is not guaranteed the complete learning of the true parameter. This is made precise in Propositions 4.3 and 4.4.

Therefore the question of complete learning depends upon whether the agent chooses an action before (model A) or after (model B) the agent observes the exogenous process. Complete learning is more likely in the latter case (model B), but even in this case further conditions are required. However these conditions take the form of checking for the nonlinearity of the solution to the one-period problem in z. Since in general the one-period problem can be solved quite easily, this condition can usually be verified. In section 4.4 we consider as an example the problem of a monopolist seeking to maximize an expected discounted sum of profits with unknown demand curves, and discuss the conditions under which we are assured that the monopolist will learn all the parameters of the demand curve in the limit.

The rest of this section is organized as follows. We first formally describe the infinite-horizon optimization problem facing the agent. We provide only a sketch here; one should consult Kiefer and Nyarko (1989) for details (especially for the measure-theoretic flourishes!). In section 4.2 we indicate that the limit points of the optimal action process solve certain one-period problems, and in section 4.3 we discuss the question of complete learning of the true parameter vector. We end in section 4.4 with an example of a monopolist maximizing an expected discounted sum of profits with unknown demand curve.

4.1. Model

Denote by $p(\mathrm{d}\varepsilon)$ the common marginal probability distribution of ε_t. On (Ω, F, P) is also defined the *exogenous* process $\{z_t\}$, that the agent observes but over which the agent has no control. We assume that $\{z_t\}$ is independent and identically distributed, with common marginal probability distribution, $q(\mathrm{d}z)$.

Let $\bar{X}$ be the action space, assumed to be a compact subset of R^1. The parameter space is given by $H = R^3$. If the true parameter value is $\Theta = (\alpha, \beta, \phi)$ and the agent chooses the action x_t in $\bar{X}$ at date t, then the agent will observe the outcome y_t given by $y_t = \alpha + \beta x_t + \phi z_t + \varepsilon_t$.

Let $P(H)$ be the set of probability distributions on H and let $\mu_0 \in P(H)$ be the prior distribution of the true parameter $\Theta = (\alpha, \beta, \phi) \in H$. If μ_{t-1} is the prior distribution at the beginning of date t, and the agent chooses an action x_t and observes z_t and y_t, then the agent updates the prior using Bayes' rule to obtain the posterior distribution, $\mu_t = \Gamma(x_t, z_t, y_t, \mu_{t-1})$. Under standard conditions the updating rule Γ is well-defined and continuous and we assume this throughout.

Let u: $\bar{X} \times R^1 \times R^1$ be the utility function; in particular, if the agent chooses action x_t and observes z_t and y_t, then the utility the agent derives is $u(x_t, z_t, y_t)$. We assume that u is bounded and continuous in its arguments.

The partial history at dates n, h_n, is the vector of past values of the (x_t, z_t, y_t) process; i.e., $h_n = ((x_1, z_1, y_1), \ldots, (x_{n-1}, z_{n-1}, y_{n-1}))$.

We consider two different models, models A and B depending upon whether the agent chooses the date t action, x_t, before or after observing the realization of z_t.

Model A. Here the agent chooses date t action x_t *before* observing the value of z_t. A policy for model A is a sequence of functions $\pi^{\mathrm{A}} = \{\pi_t^{\mathrm{A}}\}_{t=1}^{\infty}$, where for each $t \geq 1$, the policy function π_t^{A} specifies the date t action, $x_t = \pi_t^{\mathrm{A}}(h_t)$, as a Borel function of the partial history $h_t = ((x_1, z_1, y_1), \ldots, (x_{t-1}, z_{t-1}, y_{t-1}))$. Note that the action $x_t = \pi_t^{\mathrm{A}}(h_t)$ is not a function of z_t. The reward function for model A, r, is then defined by

$$r(x_t, \mu_{t-1}) = \int u(x_t, z_t, y_t) p(\mathrm{d}\varepsilon_t) q(\mathrm{d}z_t) \mu_{t-1}(\mathrm{d}\Theta), \tag{8a}$$

where $y_t = \alpha + \beta x_t + \phi z_t + \varepsilon_t$ and $\Theta = (\alpha, \beta, \phi)$.

Model B. Here the agent chooses the date t action x_t *after* observing the value of z_t. A policy of model B is a sequence of functions $\pi^{\mathrm{B}} = \{\pi_t^{\mathrm{B}}\}_{t=1}^{\infty}$, where for each $t \geq 1$, the policy function π_t^{B} specifies the date t action, $x_t = \pi_t^{\mathrm{B}}(h_t, z_t)$, as a Borel function of the partial history $h_t = ((x_1, z_1, y_1), \ldots, (x_{t-1}, z_{t-1}, y_{t-1}))$ and the observation of z_t. The reward function for model B, r, is then a function of z_t and is defined by

$$r(x_t, z_t, \mu_{t-1}) = \int u(x_t, z_t, y_t) p_t(\mathrm{d}\varepsilon_t) \mu_{t-1}(\mathrm{d}\Theta), \tag{8b}$$

where $y_t = \alpha + \beta x_t + \phi z_t + \varepsilon_t$ and $\Theta = (\alpha, \beta, \phi)$.

We assume that in both model A and model B the reward function is strictly concave[4] in x.

Let δ in $(0, 1)$ be the discount factor. Any policy π will generate a sum of expected returns equal to

$$V_{\pi}(\mu_0) = \mathrm{E} \sum_{t=1}^{\infty} \delta^{t-1} r(x_t, \mu_{t-1}) \tag{9a}$$

[4] It will become clear later on that we require this strict concavity assumption only for prior distributions that could be limiting distributions. Hence, we may ignore the case of the prior μ being concentrated on $\beta = 0$, in which case the strict concavity of the reward function r may not hold. Further we note that outside of the case where the prior has point mass on $\beta = 0$, the strict concavity of r in x is implied by (but does not imply) the strict concavity of the utility function in its arguments x and y.

for model A, and

$$V_\pi(\mu_0, z_1) = \mathrm{E}\sum_{t=1}^{\infty} \delta^{t-1} r(x_t, z_t, \mu_{t-1}) \tag{9b}$$

for model B, where the x_t and μ_t are those induced by the policy π. A policy π^* is said to be an *optimal policy* if for all policies π and all priors μ_0 in $P(H)$,

$$V_{\pi^*}(\mu_0) \geq V_\pi(\mu_0) \tag{10a}$$

for model A, and

$$V_{\pi^*}(\mu_0, z_1) \geq V_\pi(\mu_0, z_1) \quad \text{for each } z_1, \tag{10b}$$

for model B. 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. The following follows from standard dynamic programming arguments:

Theorem 4.1. For both models A and B: An optimal policy π^* exists; the value function V is continuous in its arguments; and if $\{x_t, \mu_{t-1}\}$ is the sequence of actions and posterior distributions induced by the optimal policy, π^*, then for each t, x_t solves the functional equation below, for models A and B respectively,

$$V(\mu_{t-1}) = \max_{x \text{ in } \bar{X}} r(x, \mu_{t-1}) + \delta \int V(\mu_t) p(\mathrm{d}\varepsilon_t) q(\mathrm{d}z_t) \mu_{t-1}(\mathrm{d}\Theta) \tag{11a}$$

and

$$V(\mu_{t-1}, z_t) = \max_{x \text{ in } \bar{X}} r(x, z_t, \mu_{t-1}) + \delta \int V(\mu_t, z_{t+1}) p(\mathrm{d}\varepsilon_t) \mu_{t-1}(\mathrm{d}\Theta) q(\mathrm{d}z_{t+1}), \tag{11b}$$

where $\mu_t = \Gamma(x, z_t, y_t, \mu_{t-1})$, $y_t = \alpha + \beta x + \phi z_t + \varepsilon_t$, and $\Theta = (\alpha, \beta, \phi)$.

4.2. *The convergence of the optimal action process*

We now characterize the limit points of the optimal action sequence, $\{x_t\}$. First we study model A. Let $h^0(\mu)$ be the optimal action for the static (one-period) problem when the prior distribution is μ; i.e., suppose that

$h^0(\mu)$ solves the problem

$$\max_{x \text{ in } \bar{X}} r(x, \mu). \tag{12a}$$

By assumption $r(x, \mu)$ is strictly concave in x, hence $h^0(\mu)$ is uniquely defined. Now consider a fixed sample path, and recall from Lemma 2.1 that the posterior process, μ_t, converges to some limiting distribution, μ_∞. In the lemma below we show that if x' is any limit point for the optimal action sequence $\{x_t\}$, then $x' = h^0(\mu_\infty)$. Since h^0 is uniquely defined, $\{x_t\}$ can have only one limit point, and hence x_t must converge [to $h^0(\mu_\infty)$]. [Similar results have been obtained in Kiefer and Easley (1988) and Kiefer and Nyarko (1989).]

Lemma 4.1a. On almost every sample path, the optimal action process, x_t, converges to some limit x' which satisfies $x' = h^0(\mu_\infty)$, where μ_∞ is the limiting posterior distribution and h^0 is the optimal policy function for the static (one-period) problem.

We now move to model B. Note that in model B the agent observes z_t before taking the action x_t. The static (one-period) problem will therefore depend on the prior, μ, and the observation of the $\{z_t\}$ process at the beginning of the period. Denote by $h^0(\mu, z)$ the optimal action in the static problem; i.e., the solution to the problem

$$\max_{x \text{ in } \bar{X}} r(x, z, \mu). \tag{12b}$$

Again, since we assume that $r(x, z, \mu)$ is strictly concave in x, $h^0(\mu, z)$ is uniquely defined.

We proceed to obtain the analogue of Lemma 4.1a for model B. We know that the posterior process will converge (to μ_∞). Hence, in the limit the optimal action, x_t, will depend very much on the value of z_t. Below we show that to each limit point of z_t will correspond a unique limit point of x_t. Hence in contrast to model A, since we have shown in Lemma 3.1 that the $\{z_t\}$ process does not converge, x_t will therefore not necessarily converge in model B. Further, we show in the lemma below that if we fix a sample path, and suppose that z' is any limit point of $\{z_t\}$ and x' is the corresponding limit point of x_t, then $x' = h^0(\mu_\infty, z')$; i.e., x' solves the static (one-period) problem with prior μ_∞ and the observation z' of the exogenous ($\{z_t\}$) process at the beginning of the period.

Lemma 4.1b. There exists a set of sample paths, A, with $P(A) = 1$, with the following properties. Fix a sample path in A; if there is a sub-sequence t_n such *that z_{t_n} converges to some z' as $n \to \infty$, then $\lim_{n \to \infty} x_{t_n} = x'$ exists and satisfies $x' = h^0(\mu_\infty, z')$, where μ_∞ is the limiting posterior distribution and h^0 is the optimal policy function for the static (one-period) problem.*

4.3. The question of complete learning of the true parameter vector

Lemmas 4.1a and 4.1b characterized the limit points of the optimal action sequence. We now seek to determine what the limiting posterior distribution is, and, in particular whether it is concentrated on the true parameter vector.

First we study model A. Recall that $h^0(\mu)$ is the solution to the one-period problem with prior μ [see (12a) above]. Fix a sample path and let μ_∞ be the limiting posterior distribution on that sample path. We know from Lemma 4.1a that the $\{x_t\}$ process converges to $x' = h(\mu_\infty)$. Since the exogenous process, $\{z_t\}$, is assumed i.i.d., if each z_t has at least two points in its support (i.e., z_t is not a constant), then from Lemma 3.2 above the $\{z_t\}$ process will have at least two limit points, z' and z'' (say), on each sample path. Hence applying Theorem 2.2, there exists F_∞-measurable random variables y' and y'' such that μ_∞ has support on both of the sets

$$\begin{aligned} M' &= \{(\alpha', \beta', \phi') : y' = \alpha' + \beta' x' + \phi' z'\}, \\ M'' &= \{(\alpha', \beta', \phi') : y'' = \alpha' + \beta' x' + \phi' z''\}. \end{aligned} \tag{13}$$

Further the true parameter value lies in both M' and M''. However there is only one value of ϕ' [equal to $(y' - y'')/(z' - z'')$] that simultaneously can be in both M' and M''. Also there is only one value of $(\alpha' + \beta' x')$ that can lie in both M' and M''. Hence we have shown:

Proposition 4.2. In model A, if the z_t process is not degenerate (i.e., has at least two limit points in its support), then the agent will learn the value of ϕ and will also learn a linear relationship that α and β will satisfy (i.e., there exists an F_∞-measurable random variable y''' such that both the true parameter and the support of μ_∞ lie in the set $M = \{(\alpha', \beta', \phi') : y''' = \alpha' + \beta' x'\}$).

One may ask whether we can strengthen Proposition 4.2 to obtain the complete learning of α and β. This does not seem too promising. Consider the situation where the true value of ϕ is known. Then we may write the regression equation as $m_t = \alpha + \beta x_t + \varepsilon_t$ where $m_t = y_t - \phi z_t$. This model is the same as the simple regression model, $y_t = \alpha + \beta x_t + \varepsilon_t$, for which from Kiefer and Nyarko (1989) we know that there may be incomplete learning of the parameters α and β.

We now move on to model B. The convergence of the posterior process will depend upon whether the optimal policy function for the one-period (static) problem, $h^0(\mu, z)$, is linear or nonlinear in z. Suppose first that h^0 is linear in z (for fixed μ). In particular fix a sample path, let μ_∞ be the limiting posterior distribution on that path, and suppose that $h^0(\mu_\infty, z) = c + dz$, where c and d may be functions of μ_∞, but independent of z. Then in very much the same way that Proposition 4.1 above was proved we can show:

Proposition 4.3. *In model B, suppose that the z_t process is not degenerate (i.e., has at least two points in its support). Then there exists a set of sample paths, A, where $P(A) = 1$ with the following property: For fixed sample path in A if μ_∞ is the limiting posterior distribution and $h^0(\mu_\infty, z) = c + dz$, then the agent will learn the values of $(\alpha + \beta c)$ and $(\beta d + \phi)$ [i.e., there exists F_∞-measurable random variables $\hat{y}$ and $\hat{y}'$ such that the true parameter vector and the support of μ_∞ lie in the set $M = \{(\alpha', \beta', \phi')\colon\ \hat{y} = \alpha' + \beta' c \text{ and } \hat{y}' = \beta' d + \phi\}$].*

Again one may ask whether Proposition 4.3 may be strengthened to obtain the complete learning of the true parameter vector. Again this is not possible in general. Let μ be the prior distribution and suppose that the one-period optimal action given observation, z, of the exogenous process is $h^0(\mu, z) = c + dz$, and assume that to begin with the agent knows the values of $(\alpha + \beta c)$ and $(\beta d + \phi)$. From the example of Kiefer and Nyarko (1989) the agent may choose for the infinite horizon the action $x = h^0(\mu, z)$, which is the one-period optimal action, if the discount factor is sufficiently low. However, the choice of such an action results in a regression equation $y = (\alpha + \beta c) + (\beta d + \phi)z + \varepsilon$, which yields no new information to the agent. The agent will therefore not learn the true parameter vector.

Next we study the situation in which the optimal policy function for the static (one-period) problem is nonlinear. We require the following nonlinearity restriction (which requires that for fixed μ, h^0 has no linear portions):

Nonlinearity Assumption. *Fix a prior distribution μ. Then there does not exist a z' and z'' (with z' different from z'') and a number m in $(0,1)$ such that*

$$h^0(\mu, mz' + (1-m)z'') = mh^0(\mu, z') + (1-m)h^0(\mu, z''). \tag{14}$$

Under the above condition, if we assume that the support of the z_t process contains at least three points, then we may very easily show that the agent will learn all the parameters, α, β, and ϕ, of the regression equation. We proceed to prove such a result.

Proposition 4.4. *In model B, if the z_t process has at least three points in its support and the optimal policy function h^0 satisfies the nonlinearity assumption above, then the agent will learn the true values of the parameters α, β, and ϕ (i.e., the limiting posterior distribution, μ_∞, will be concentrated on the true parameter values of α, β, and ϕ).*

Propositions 4.2–4.4 give us information on the convergence of the posterior process in terms of the linearity or otherwise of the one-period optimal policy functions. The usefulness of these results is of course due to the fact that in most cases the one-period problems can be very easily solved and the linearity of the optimal policy functions determined. Below we illustrate the results above with an example of a monopolist with unknown demand curve.

4.4. Example: Monopolist with unknown demand curve

Suppose that at each date a monopolist has to set a price p_t of a single good that the monopolist sells. Given the chosen price, p_t, the quantity of the good, q_t, that the public purchases is given by the demand curve

$$q_t = \alpha + \beta p_t + \phi z_t + \varepsilon_t, \tag{15}$$

where z_t is an i.i.d. exogenous process and ε_t is an i.i.d. noise term with mean zero and bounded variance that the monopolist does not observe. The parameters α, β, and ϕ are unknown to the monopolist and μ is the monopolist's prior over these parameters. The cost of producing output q is $C(q)$, so the monopolist's profit at date t is $\pi_t = p_t q_t - C(q_t)$. The objective of the monopolist is to maximize the expected sum of discounted utilities of profit, $\mathrm{E}\sum_{t=1}^{\infty}\delta^{t-1}u(\pi_t)$. We seek to discuss the question of whether the monopolist will over time learn the parameters of the demand curve.

First, from section 4.3 above we know that if the monopolist has to announce date t prices before observing z_t (this is model A), then it is possible that the monopolist will not learn the true parameter vector. If alternatively the monopolist observes z_t before choosing the price P_t, then complete learning of the true parameter vector depends upon the linearity of the optimal one-period price $p = h^0(\mu, z)$, as a function of the observation of z, with incomplete learning if this function is linear and complete learning if the function is nonlinear [i.e., satisfies the nonlinearity condition (14)]. We now determine the linearity of the one-period optimal policy function, with different specifications of the cost function and the utility function. All of these cases are for the situation where the monopolist observes the z_t before choosing the price p_t (i.e., model B).

Case I: Risk-neutral monopolist with zero costs. The expected profits in this case will be given by

$$\begin{aligned}\mathrm{E}\pi_t &= \mathrm{E}[p_t(\alpha + \beta p_t + \phi z_t + \varepsilon_t)] \\ &= p_t\,\mathrm{E}\alpha + p_t^2\,\mathrm{E}\beta + p_t z_t\,\mathrm{E}\phi.\end{aligned} \tag{16}$$

Assuming that $E\beta < 0$, the above function of p_t will be maximized at $p_t = -(E\alpha + z_t E\phi)/2E\beta$. The optimal policy function for the one-period problem is then given by

$$h^0(\mu, z) = -(E_\mu \alpha + z E_\mu \phi)/2E_\mu \beta. \tag{17}$$

This is a linear function z, for fixed prior μ. In this case Proposition 4.3 holds and we are not guaranteed that the agent learns the true parameter vector.[5]

Case II: Risk-neutral monopolist with nonlinear cost function $C(q)$. If $C(q)$ is sufficiently nonlinear, then it is clear that the optimal policy function for the one-period problem will also be nonlinear and indeed satisfy the nonlinearity assumption required in the proof of Proposition 4.4. In that case, from Proposition 4.4, there will be complete learning of the true parameter. We note however that the quadratic cost function will not work; it results in a linear optimal policy function for the one-period problem. However, practically any other strictly convex cost function will result in a nonlinear optimal policy function for the static problem, and therefore complete learning of all the parameters of the demand curve.

Case III: Risk-averse monopolist. Alternatively suppose that the monopolist is risk-averse and seeks to maximize the expected utility of profits, $Eu(\pi)$, where u is a strictly concave function. Then, in general, the optimal policy function for the one-period will be nonlinear and, in particular, satisfy the nonlinearity condition. We may therefore apply Proposition 4.4 to conclude that a monopolist solving an infinite-horizon problem in this case will over time learn the values of all the parameters of the demand curve.

5. Conclusion

In this paper we have provided a technique, counting the number of 'equations' and 'unknowns', that is useful in determining when the Bayesian posterior process of a linear regression model will converge to point mass on the true parameter values. This technique is illustrated with two examples – i.i.d. regressors and lagged dependent regressors. We then study the optimal control of a linear regression process. We indicate that the question of complete learning is answered by looking at processes resulting from the use of myopic or one-period optimal policies and then counting the number of 'equations' and 'unknowns'. We show that complete learning of all

[5] However, Proposition 4.3 tells us that the monopolist will learn the values of $(\alpha - \beta(E\alpha/E\beta))$ and $(1 - \beta(E\phi/E\beta))$ where the expectations are those with respect to the limiting posterior distribution, μ_∞, which obviously will be known to the agent in the limit.

the parameters of the regression process is likely when there is an exogenous process (say GNP) that agents observe before taking their actions. The results are illustrated with the example of a monopolist seeking to maximize an expected sum of discounted profits with an unknown demand curve.

Appendix

Proof of Lemma 2.1. This follows almost immediately from observing that, for any set A, the sequence $z_t = P[\Theta \in A | F_t]$ is a martingale sequence with respect to F_t, so one may apply the martingale convergence theorem. [See, e.g., Kiefer and Nyarko (1989, theorem 4.1).]

Proof of Theorem 2.2. First we state the following lemma due to Taylor (1974).

Lemma A.1. Let $\{v_t\}$ be a sequence of independent random variables with zero mean and uniformly bounded second moment. Let $\{z_t\}$ be a sequence of random variables such that for each t and t', with $t < t'$, $v_{t'}$ is independent of $\{z_1, z_2, \ldots, z_t\}$; then for almost every realization such that $\sum_{t=1}^{T} z_t^2 \to \infty$,

$$\lim_{T\to\infty} \sum_{t=1}^{T} z_t v_t \Big/ \sum_{t=1}^{T} z_t^2 = 0. \tag{18}$$

Proof. One applies Taylor (1974, lemmas 1–3) with minor modifications. ∎

We use Lemma A.1 above to prove Lemma A.2 below which is key in proving the theorem. Let X_t denote the vector $(x_{1t}, \ldots, x_{kt})$. Given any two vectors a and b in R^k, we use notation $a < b$ if $a_i < b_i$ for each coordinate $i = 1, \ldots, k$. A vector a in R^k will be called a rational vector if for each coordinate $i = 1, \ldots, k$, a_i is a rational number. Let $1_{\{w \in K\}}$ be the indicator function on K, i.e., $1_{\{w \in K\}}$ is 1 if $\{w \in K\}$ and is zero otherwise. We now state Lemma A.2.

Lemma A.2. There exists a set A of sample paths with $P(A) = 1$ with the following property: For each fixed sample path in A and for any two fixed *rational vectors a and b, with* $a < b$, *if* $\sum_{t=1}^{\infty} 1_{\{a < X_t < b\}} = \infty$, *then*

$$\lim_{T \to \infty} \sum_{t=1}^{T} \varepsilon_t 1_{\{a < X_t < b\}} \bigg/ \sum_{t=1}^{T} 1_{\{a < X_t < b\}} = 0. \tag{19}$$

Proof. Using Lemma A.1, it is immediate that for fixed vectors $a < b$, (19) holds on some set $A(a, b)$ with $P(A(a, b)) = 1$. Define A to be the intersection over all rational vectors $a < b$, then A satisfies all the conclusions of the lemma. ■

We now prove Theorem 2.2. Let $\{t_n\}$ and $X' = (x'_1, \ldots, x'_k)$ be as in Theorem 2.2. Below we define y', and hence the set M of the theorem, and show that on the set A of Lemma A.2, $\Theta = (\alpha, \beta_1, \ldots, \beta_k)$ lies in M. This will conclude the proof of Theorem 2.2, for in that case $1_{\{\Theta \in M\}} = 1$ with probability one, hence

$$\mu_\infty(M) = \mathrm{E}\left[1_{\{\Theta \in M\}} | F_\infty\right] = \mathrm{E}\left[1 | F_\infty\right] = 1 \quad \text{a.e.} \tag{20}$$

The true parameter vector satisfies the regression equation

$$y_t = \alpha + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + \varepsilon_t. \tag{21}$$

Let a and b be any two rational vectors in R^k such that $a < X' < b$ where $X' = (x'_1, \ldots, x'_k)$ (i.e., for each coordinate i, $a_i < x'_i < b_i$). If for expositional convenience we write $1_t = 1_{\{a < X_t < b\}}$, then multiplying both sides of (21) by 1_t, summing over t and then dividing by $\sum 1_t$, one obtains

$$\bar{y}_T = \alpha + \beta_1 \bar{x}_{1T} + \cdots + \beta_k \bar{x}_{kT} + \bar{\varepsilon}_T, \tag{22}$$

where if $S(T) = \sum_{t=1}^{T} 1_t$, $\bar{y}_T = \sum_{t=1}^{T} y_t 1_t / S(T)$, $\bar{x}_{iT} = \sum_{t=1}^{T} x_{it} 1_t / S(T)$ for each i, and $\bar{\varepsilon}_T = \sum_{t=1}^{T} \varepsilon_t 1_t / S(T)$. From the definition of X' as the limit of the vector X_{t_n}, for each i, $a_i < x_{it_n} < b_i$ for all but at most finitely many n; hence $\sum_{t=1}^{\infty} 1_{\{a < X_t < b\}} = \infty$, (19) holds and $\bar{\varepsilon}_T \to 0$. Taking the limit superior as $T \to \infty$ on both sides of (22), results in

$$\limsup_{T \to \infty} \bar{y}_T = \limsup_{T \to \infty} \left(\alpha + \beta_1 \bar{x}_{1T} + \cdots + \beta_k \bar{x}_{kT}\right). \tag{23}$$

Choose a sequence of pairs of rational vectors, $\{a^m, b^m\}_{m=1}^{\infty}$, such that for each m, $a^m < X' < b^m$, and as $m \to \infty$, $a^m \to X'$ and $b^m \to X'$. For fixed m, let $\bar{y}_T^m$ and $\bar{x}_{iT}^m$ be the same as $\bar{y}_T$ and $\bar{x}_{iT}$ defined above except that the vectors a^m and b^m replace the vectors a and b. Then we obtain the equivalent of (23) for each m:

$$\limsup_{T\to\infty} \bar{y}_T^m = \limsup_{T\to\infty} \left(\alpha + \beta_1 \bar{x}_{1T}^m + \cdots + \beta_k \bar{x}_{kT}^m\right). \tag{24}$$

Next we take limits as $m \to \infty$ on both sides of (24) above. It is easy to show by a simple $\varepsilon - \delta$ argument that

$$\lim_{m\to\infty} \limsup_{T\to\infty} \left(\alpha + \beta_1 \bar{x}_{1T}^m + \cdots + \beta_k \bar{x}_{kT}^m\right)$$

$$= \alpha + \beta_1 x'_{1T} + \cdots + \beta_k x'_{kT}. \tag{25}$$

Hence defining $y' = \lim_{m\to\infty} \limsup_{T\to\infty} \bar{y}_T^m$, we obtain

$$y' = \alpha + \beta_1 x'_1 + \cdots + \beta_k x'_k. \tag{26}$$

With y' defined, (26) implies that $\Theta = (\alpha, \beta_1, \ldots, \beta_k)$ lies in M. This concludes[6] the proof of Theorem 2.2. Q.E.D.

Proof of Proposition 3.1. From condition (ii) the vector $(1, X_t) = (1, x_{1t}, \ldots, x_{kt})$ has at least $k+1$ linearly independent vectors in its support. Let us denote these vectors by $(1, b^1), (1, b^2), \ldots, (1, b^{k+1})$, where for each $j = 1, \ldots, k+1$, b^j is a vector in R^k. If X_t is discrete, then we may assume that $P(X_t = b^j) > 0$ for each $j = 1, \ldots, k+1$. Otherwise, let B^j be a neighborhood in R^k around b^j. We choose these neighborhoods sufficiently small (i.e., sufficiently close to the b^j's) so that if $\{a^1, \ldots, a^{k+1}\}$ is any collection of $k+1$ vectors in R^k, and for each j a^j lies in B^j, then the $k+1$ vectors $\{(1, a^1), \ldots, (1, a^{k+1})\}$ are linearly independent. (It is easy to show that this is always possible.) Note that since b^j is in the support of X_t, $P(X_t$ lies in $B^j) > 0$.

Fix any $j = 1, \ldots, k+1$. From Lemma 3.2 above, outside of a null set X_t has a limit point in B^j; let us call this limit point $X^j = (x_1^j, \ldots, x_k^j)$. From Theorem 2.2 above the agent will learn the value of $\alpha + \beta_1 x_1^j + \cdots + \beta_k x_k^j$. Since, by construction, the $k+1$ vectors $(1, X^1), \ldots, (1, X^{k+1})$ are linearly independent, the agent will necessarily learn the true parameter vector.

Q.E.D.

[6]Note that we have overlooked the F_∞-measurability of y'. However, observe that y' is a limit of variables that are each in F_∞. To complete the verification that y' is F_∞-measurable we need to show that the rational vectors a, b and a^m, b^m, for each m, can be chosen as a function of the sample path in a measurable way. As we mentioned in a footnote to Theorem 2.2, we assume that the limit points x'_i are chosen measurably. If this is done, then one can apply the same technique as in Kiefer and Nyarko (1989, p. 584, in the proof of lemma 3.4, where a function $h(x, x')$ is used to measurably choose the rational vectors) to conclude that y' is indeed F_∞-measurable.

Proof of Proposition 3.3. We do not assume that the support of the distribution of the error term, ε, has bounded support. This results in the proof being much longer than otherwise. We prove this version as we believe it may be of independent interest.

From the discussion preceding Proposition 3.3, it suffices to show that the $\{y_t\}$ process has at least two finite limit points. One can show by a simple induction argument that

$$y_t = \beta^t y_0 + \alpha \sum_{i=1}^{t} \beta^{t-i} + \sum_{i=1}^{t} \beta^{t-i} \varepsilon_t. \tag{27}$$

The first two terms on the right-hand side of (27) above converge to finite limits as $t \to \infty$. Hence y_t will have two finite limit points if $S_t = \sum_{i=1}^{t} \beta^{t-i} \varepsilon_t$ has two finite limit points, which we now proceed to show. It must be stressed that we are not assuming that the support of the $\{\varepsilon_t\}$ is bounded. Hence S_t may have limit points at ∞ or $-\infty$ (indeed, if the support of ε_t is the real line, then one can show that $\limsup S_t = \infty$ and $\liminf S_t = -\infty$, a.e.).

We now summarize how the proof will proceed. First we show that $\liminf S_t^2 < \infty$; then we show that $\liminf S_t^2$ is actually a constant, equal to some k^2 (say). This means that on almost very sample path S_t has a limit point at either k or $-k$ (with k independent of the sample path). Next we show that if S_t has one limit point, at k say, then it also has another limit point. This will then conclude the proof.

One can show very easily that

$$\mathrm{E}\, S_t^2 = \mathrm{E}\, \varepsilon_1^2 \sum_{i=1}^{t} \beta^{2(t-i)} \le \mathrm{E}\, \varepsilon_1^2/(1-\beta^2). \tag{28}$$

Then using Fatou's lemma [see, e.g., Chung (1974, p. 42)],

$$\mathrm{E} \liminf S_t^2 \le \liminf \mathrm{E}\, S_t^2 \le \mathrm{E}\, \varepsilon_1^2/(1-\beta^2) < \infty, \tag{29}$$

which implies that $\liminf S_t^2 < \infty$, a.e.

Next fix an integer N. Write, for $t > N$,

$$S_t = \beta^t \sum_{i=1}^{N} \beta^{-i} \varepsilon_i + \sum_{i=N+1}^{t} \beta^{t-i} \varepsilon_i. \tag{30}$$

Let $H'_N = \sigma(\{\varepsilon_{N+1}, \varepsilon_{N+2}, \dots\})$, the σ-algebra generated by the random variables $\{\varepsilon_{N+1}, \varepsilon_{N+2}, \dots\}$. Note that as $t \to \infty$ the first summation in (30) above tends to zero. Hence the limit properties of S_t depend only on $\{\varepsilon_{N+1}, \varepsilon_{N+2}, \dots\}$, and hence $\liminf S_t^2$ is measurable with respect to H'_N.

Since N is arbitrary, $\liminf S_t^2$ is therefore measurable with respect to the tail σ-algebra $H'_\infty = \bigcap_{N=1}^{\infty} H'_N$. Since the $\{\varepsilon_t\}$ are independent, by Kolmogorov's 0–1 law [see, e.g., Chung (1974, theorem 8.1.4)] each event in H'_∞ has probability zero or one. Hence $\liminf S_t^2$ must be a constant (a.e.). [To see this let $Z = \liminf S_t^2$ and note that if Z were not a constant, then there exists a number d such that $P(Z \leq d) > 0$ and $P(Z > 0) > 0$; since $\{Z \leq d\}$ and $\{Z > 0\}$ are both H'_∞-measurable, by the Kolmogorov 0–1 law $P(Z \leq d) = 1 = P(Z > d)$, a contradiction.] We may therefore write $\liminf S_t^2 = k^2$ for some $k \geq 0$.

We have therefore shown that S_t has a limit point at either k or $-k$. Define $A^+ = \{S_t$ has limit point at $k\}$ and $A^- = \{S_t$ has limit point at $-k\}$; then $P(A^+ \cup A^-) = 1$. We show in Claim 2 that on A^+, S_t has a limit point different from k; similar arguments show that on A^- S_t has a limit point different from $-k$. Hence S_t has two limit points on each sample path, and this will conclude the proof of the proposition.

Under the assumption that the ε_t process is not degenerate and has mean zero, there exists e and e' such that $0 < e < e' < \infty$ and $P(-e' \leq \varepsilon_1 \leq -e) > 0$. We can show:

Claim 1. Suppose S_t lies in $G = [-(k + e/2), k + e/2]$ and ε_{t+1} lies in $[-e', -e]$. Then S_{t+1} lies in $G' = [-(k + e/2) - e', k - e/2]$.

Proof. Let G and G' be as defined in the claim. Suppose $S_t \in G$; then $|\beta| < 1$ implies that $\beta S_t \in G$, so when ε_{t+1} lies in $[-e', -e]$, $S_{t+1} = \beta S_t + \varepsilon_{t+1}$ lies in $[-(k + e/2) - e', k + e/2 - e] = G'$. ■

Next we prove in Claim 2 below that on almost every sample path in A^+, S_t visits the set G' defined in Claim 1 above infinitely often and hence has a limit point in G', which is necessarily different from k. By obvious modifications one obtains similarly that on A^-, S_t also has two limit points. Since $P(A^+ \cup A^-) = 1$, this concludes the proof of Proposition 3.3.

Claim 2. On almost every sample path in A^+, S_t lies in the set G' (defined in Claim 1 above) infinitely often.

Proof. Let $H_n = \sigma(\{\varepsilon_1, \ldots, \varepsilon_n\})$, the σ-algebra induced by the random variables $\{\varepsilon_1, \ldots, \varepsilon_n\}$. Define $A_t = \{S_t \in G\}$ (with G as in Claim 1), $B_t = \{\varepsilon_t$ lies in $[-e', -e]\}$, and $C_t = \{S_t$ lies in $G'\}$, and note that for each t, A_t, B_t, and C_t all are H_t-measurable. From Claim 1 we know that conditional on A_t, the probability that C_{t+1} occurs is at least $P(-e' \leq \varepsilon_{t+1} \leq -e) = P(-e' \leq \varepsilon_1 \leq -e) > 0$. But recall that from the definition of A^+, $S_t \in A_t$ infinitely often on A^+, so $\sum_{t=1}^{\infty} P(C_{t+1} | H_t) = \infty$ on A^+. Therefore from the 'conditional' Borel–Cantelli-Levy lemma [see, e.g., Chow, Robbins, and Siegmund (1971,

p. 26)], we conclude that except for a set with probability zero, C_t occurs infinitely often A^+. ∎

Proof of Lemma 4.1a. In Lemma 4.1b a similar result is stated for model B; one may easily construct a proof of Lemma 4.1a from the proof of Lemma 4.1b which is presented below, so we omit the proof.

Proof of Lemma 4.1b. From the functional equation (11b) there is a set A with $P(A)=1$ such that on A, for all t (simultaneously),

$$V(\mu_{t-1}, z_t) = r(x_t, z_t, \mu_{t-1}) + \delta \int \mathrm{E}[V(\mu_t, z_{t+1}) \mid F_{t-1}] q(\mathrm{d}z_{t+1}), \tag{31}$$

where $y_t = \alpha + \beta x_t + \phi z_t + \varepsilon_t$, $\mu_t = \Gamma(x_t, z_t, y_t, \mu_{t-1})$; the E operator is taken over (α, β, ϕ) and ε_t using μ_{t-1} and $p(\mathrm{d}\varepsilon)$; and where q is the common marginal distribution of z_t. Next fix a sample path in A. Let t_n and z' be as stated in the lemma, and let x' be any limit point of the sequence $\{x_{t_n}\}_{n=1}^{\infty}$. We shall show that $x' = h^0(\mu_\infty, z')$ where μ_∞ is the limiting posterior distribution (for the fixed sample path in A); since $r(x, z, \mu_\infty)$ is assumed strictly concave in x, h^0 and therefore x' is unique, and this would conclude the proof of the lemma.

Replace t with t_n in (31) and take limits as $n \to \infty$; then noting that r, Γ, and V are continuous in their arguments and V is bounded, one can show[7]

$$V(\mu_\infty, z') = r(x', z', \mu_\infty) + \delta \int \mathrm{E}[V(\mu'_\infty, z) \mid F_\infty] q(\mathrm{d}z), \tag{32}$$

where $y = \alpha + \beta x' + \phi z' + \varepsilon$, $\mu'_\infty = \Gamma(x', z', y, \mu_\infty)$, and the expectation is taken over (α, β, ϕ) and ε with respect to μ_∞ and $p(\mathrm{d}\varepsilon)$, respectively.

Now we show that x' solves the problem

$$\max_{x \text{ in } \bar{X}} r(x, z', \mu_\infty). \tag{33}$$

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

$$r(x', z', \mu_\infty) < r(\hat{x}, z', \mu_\infty). \tag{34}$$

[7] Indeed one uses Chung (1974, theorem 9.4.8) to conclude that $\mathrm{E}[V(\mu_{t_n}, z) | F_{t_n - 1}] \to \mathrm{E}[V(\mu'_\infty, z) | F_\infty]$ for each z, then one applies the Dominated Convergence Theorem.

From Theorem 2.2 there exists an F_∞-measurable random variable, y', such that μ_∞ has support on the set $M = \{(\alpha', \beta', \phi')\colon\ y' = \alpha' + \beta' x' + \phi' z'\}$. Hence, if the action x' is chosen, then under μ_∞ the regression equation becomes $y = \alpha + \beta x' + \phi z' + \varepsilon = y' + \varepsilon$. The observation, y, in the next period is then equal to y' plus a noise term so the experiment 'observe y' has no information content. Choosing any other action, $\hat{x}$, will therefore be more informative than the action x', hence from Blackwell's theorem [see, e.g., Blackwell (1951) or Kihlstrom (1984)] for each z,

$$\mathrm{E}[V(\mu'_\infty, z) \mid F_\infty] \leq \mathrm{E}[V(\hat{\mu}'_\infty, z) \mid F_\infty], \tag{35}$$

where $\hat{y} = \alpha + \beta \hat{x} + \phi z' + \varepsilon$, $\hat{\mu}'_\infty = \Gamma(\hat{x}', z', \hat{y}, \mu_\infty)$, and the E operator is the expectation taken over ε and (α, β, ϕ) with respect to μ_∞.

Putting (34) and (35) into (32) then results in

$$V(\mu_\infty, z') < r(\hat{x}, \mu_\infty, z') + \delta \int \mathrm{E}[V(\hat{\mu}'_\infty, z) \mid F_\infty] q(\mathrm{d}z), \tag{36}$$

which is a contradiction to the functional equation (11b). This proves that x' solves the problem in (36) and concludes the proof of Lemma 4.1b. Q.E.D.

Proof of Proposition 4.4. If z_t has at least three limit points in its support, then from Lemma 3.1 on almost every sample path there will be at least three limit points, z, z', and z'' (say). From Lemma 4.1b, corresponding to each limit point of the z_t process, z, z', and z'', there will be a unique limit point of the x process, x, x', and x'', respectively. One then uses Theorem 2.2 to show that the agent will learn three equations that the true parameter vector satisfies; finally one uses the nonlinearity condition to show that these three equations are linearly independent. Hence the agent necessarily learns the true parameter vector. Q.E.D.

References

Aghion, P., P. Bolton, and B. Julien, 1986, Dynamic price and quantity setting by a monopolist facing unknown demand, Manuscript (Harvard University, Cambridge, MA).

Anderson, T.W. and J.B. Taylor, 1979, Strong consistency of least squares estimates in dynamic models, Annals of Statistics 7, 484–489.

Blackwell, D., 1951, The comparison of experiments, in: Proceedings of the second Berkeley symposium on mathematical statistics and probability (University of California Press, Los Angeles, CA) 93–102.

Blume, L. and D. Easley, 1984, Rational expectations equilibrium: An alternative approach, Journal of Economic Theory 34, 116–129.

Bray, M. and N. Savin, 1986, Rational expectations equilibria, learning, and model specification, Econometrica 54, 1129–1160.

Chow, Y.S., H. Robbins, and D. Siegmund, 1971, Great expectations: The theory of optimal stopping (Houghton Mifflin, Boston, MA).

Chung, K.L., 1974, A course in probability theory (Academic Press, New York, NY).
Easley, D. and N. Kiefer, 1988, Controlling a stochastic process with unknown parameters, Econometrica 56, 1045–1064.
Feldman, M., 1987, An example of convergence to rational expectations with heterogeneous beliefs, International Economic Review 28, 635–650.
Frydman, R., 1982, Towards an understanding of market processes: Individual expectations and convergence to rational expectations equilibrium, American Economic Review 72, 652–668.
Grossman, S., R.E. Kihlstrom, and L.J. Mirman, 1977, A Bayesian approach to the production of information and learning by doing, Review of Economic Studies 44, 533–547.
Kiefer, N.M. and Y. Nyarko, 1989, Optimal control of an unknown linear process with learning, International Economic Review 30, 571–586.
Kihlstrom, R., 1984, A 'Bayesian' exposition of Blackwell's theorem on the comparison of experiments, in: M. Boyer and R.E. Kihlstrom, eds., Bayesian models of economic theory (Elsevier Science Publishers, Amsterdam).
Kihlstrom, R.E., L.J. Mirman, and A. Postlewaite, 1984, Experimental consumption and the 'Rothschild effect', in: M. Boyer and R.E. Kihlstrom, eds., Bayesian models of economic theory (Elsevier Science Publishers, Amsterdam).
Marcet, A. and T. Sargent, 1989, Convergence of least squares learning in self referential linear stochastic models, Journal of Economic Theory 48, 337–368.
McLennan, A., 1984, Price dispersion and incomplete learning in the long run, Journal of Economic Dynamics and Control 7, 331–347.
McLennan, A., 1986, Optimization with learning, Manuscript (Cornell University, Ithaca, NY).
Prescott, E.C., 1972, The multi-period control problem under uncertainty, Econometrica 40, 1043–1057.
Rothschild, M., 1974, A two-armed bandit theory of market pricing, Journal of Economic Theory 9, 185–202.
Taylor, J.B., 1974, Asymptotic properties of multiperiod control rules in the linear regression model, International Economic Review 15, 472–484.
White, H., 1984, Asymptotic theory for econometricians (Academic Press, New York, NY).