Murphy, J.L. 1966: Effects of the threat of losses on duopoly bargaining. *Quarterly Journal of Economics*, 80, 296–313.

Nyarko, Y. 1990: Learning in mis-specified models and the possibility of cycles. Economic Research Report 90-03, C.V. Starr Center for Applied Economics, New York University.

Plott, C.R. 1982: Industrial organization theory and experimental economics. *Journal of Economic Literature*, 20, 1485–1527.

Prescott, E.C. 1972: The multiperiod control problem under uncertainty. *Econometrica*, 40 (6), 1043–58.

Robson, A. 1986: The existence of Nash equilibria in reaction functions for dynamic models of oligopoly. *International Economic Review*, 27, 539–44.

Rothschild, M. 1974: A two-armed bandit theory of market prices. *Journal of Economic Theory*, 9, 185–202.

Rubinstein, A. 1979: Equilibrium in supergames with the overtaking criterion. *Journal of Economic Theory*, 21, 1–9.

Selten, R. 1965: Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragetägheit. *Zeitschrift für die Gesamte Staatswissenschaften*, 121, 301–24, 667–89.

Selten, R. 1967: Ein Oligopolexperiment mit Preisvariation und Investition. In H. Sauermann (ed.), *Beiträge zur Experimentellen Wirtschaftsforschung*, Tübingen: Mohr, 103–35.

Shapley, L. 1964: Some topics in two-person games. In M. Dresher et al. (eds), *Advances in Game Theory*, Princeton, NJ: Princeton University Press.

Sternberg, S. 1963: Stochastic learning theory. In R.D. Luce, R.R. Bush and E. Galanter (eds), *Handbook of Mathematical Psychology*, New York: Wiley.

Stoecker, R. 1980: *Wirtschaftstheoretische Entscheidungsforchung*, vol. 4: *Experimentelle Untersuchung des Entscheidungsverhaltens im Bertand-Oligopole.* Institut für Mathematische Wirtschaftsforschung, Universität Bielefeld.

Urbano, A. 1992: Monopoly experimentation: a generalization. *Revista Española de Economia*, 9 (2), 275–305.

Williams, S. 1985: Necessary and sufficient conditions for the existence of a locally stable message process. *Journal of Economic Theory*, 35, 127–54.

Woodford, M. 1990. Learning to believe in sunspots. *Econometrica*, 58, 277–307.

Zellner, A. 1971: *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.

6

Speed of Convergence of Recursive Least Squares: Learning with Autoregressive Moving-Average Perceptions

Albert Marcet and Thomas J. Sargent

Introduction

This chapter studies the convergence to a limited information rational expectations equilibrium of a self-referential system in which agents are learning by recursively updating their estimates of an autoregressive moving-average (ARMA) model for endogenous variables. In the existing literature on least squares learning, e.g. Bray (1982, 1983), Bray and Savin (1986), Fourgeaud et al. (1986), and Marcet and Sargent (1989a, b), agents are assumed to learn by recursively fitting finite order pure autoregressions. In models with private information and/or hidden state variables, the restriction to a finite order autoregressive scheme is limiting because the stochastic structure of the rational expectations equilibrium gives agents an incentive to condition on the infinite past of the variables that they observe (see Marcet and Sargent (1989b) and Sargent (1991) for elaborations of this point). It is natural to seek what in earlier work (Sargent, 1991) we called a *full order equilibrium*, namely, an equilibrium in which agents' forecasting rules achieve the minimum possible one-step-ahead forecasting error variance given the infinite record of past observations. In Sargent (1991), we described how in the context of Townsend's (1983) model[1] such an equilibrium can be supported with finite-order parameterizations by specifying that agents forecast by using ARMA schemes. Sargent (1991) studied how to formulate and compute such an equilibrium, but did not analyze convergence to it via least squares.

To study least squares learning in a simple version of such a setting, this chapter analyzes a modification of the hyperinflation model studied by Fourgeaud et al. (1986). We alter their model in just one significant way: we assume that agents do not observe the money supply, and that the only information on which they can base forecasts of future prices is current and past prices.[2] For this setup, there may exist a *limited information rational expectations equilibrium* in which the price level is a first-order ARMA process. We study whether we can expect convergence to this equilibrium by a system in which agents forecast by fitting a first-order ARMA process to prices each period, updating their estimates of the ARMA parameters recursively. We study the convergence of the resulting system under two distinct recursive algorithms for estimating ARMA processes: (i) pseudo-linear regression, and (ii) the recursive prediction error method.[3]

We study convergence by adapting arguments described by Marcet and Sargent (1989a), which in turn are based on arguments of Ljung (1977) and Ljung and Söderström (1983).[4] The ordinary differential equations governing pseudo linear regression and the recursive prediction error method are shown to differ, but to share a common rest point (the limited information rational expectations equilibrium).

The eigenvalues of the associated ordinary differential equations at the fixed point shed some light on the speeds of convergence of our two algorithms.[5] In particular, we use recent theoretical results of Benveniste et al. (1990) to get some results on rates of convergence and how they depend on those eigenvalues. We use a method for estimating the rate of convergence via simulation for situations in which we are without analytical results.

Systems in which agents form perceptions in the form of ARMA processes arise naturally in a variety of contexts. In addition to the models with private information and hidden state variables described by Marcet and Sargent (1989b) and Sargent (1991),[6] they arise in linear models with sunspots and multiple equilibria. Evans and Honkapohja (1990) describe a setup in which there are multiple equilibria differing among one another in the number of parameters in their ARMA representations. Evans and Honkapohja study the stability of these alternative equilibria in the face of some version of least squares learning. For technical reasons, Evans and Honkapohja have yet to complete their analysis of stability for the case in which the equilibria are of ARMA, as opposed to just AR, form. The results in the present chapter will be useful in contexts like theirs.

1 The Model

We adapt the inflation model of Fourgeaud et al. (1986) as follows.[7] Let y_t be the log of the price level and x_t be the log of the money supply at t. The variables (y_t, x_t) are determined by

$$y_t = \lambda E^*(y_{t+1}|y_t, w_t) + x_t + v_t \tag{6.1}$$

$$x_t = \rho x_{t-1} + u_t + d u_{t-1} \tag{6.2}$$

where λ, ρ, and d are all less than unity in absolute value, and (u_t, v_t) is a pair of mutually orthogonal white noises with variances σ_u^2 and σ_v^2 respectively. Equation (6.1) is a version of a demand function for money, while equation (6.2) is the assumed stochastic process for the money supply, which is an exogenous first-order ARMA process. In (6.1), $E^*(y_{t+1}|y_t, w_t)$ is agents' forecast of y_{t+1} at time t. Let this be given by

$$E^*(y_{t+1}|y_t, w_t) \equiv E_t(y_{t+1}) = a y_t + c w_t \tag{6.3}$$

where $|a| < 1$, $|c| < 1$. The parameters a, c and the variate w_t are determined by agents' perceptions of an ARMA(1,1) model

$$y_{t+1} = a y_t + w_{t+1} + c w_t \tag{6.4}$$

where w_t is believed to be the innovation in y_t relative to the information set $y^t = (y_t, y_{t-1}, \ldots)$. Agents assume the time-invariant model (6.4) for y_t and estimate it via a procedure to be described below. The force of (6.3) and (6.4) is that agents do not observe x_t, v_t, or u_t in (6.1) and (6.2), but do observe the record $y^t = (y_t, y_{t-1}, \ldots)$. From the perspective of agents, there are "hidden state variables."

We shall now describe how to formulate and compute an appropriate notion of a rational expectations equilibrium for this model. We do so by describing the mapping from a perceived to an actual law of motion for prices, the same sort of mapping that was utilized extensively by Marcet and Sargent (1989a) and Sargent (1991). The method is first to define the state of the system and to find its law of motion. Then we deduce the univariate law of motion for the log price level by "conditioning down," i.e. finding the projection of prices on past prices and the innovation in prices implied by the law of motion for the state of the system. This procedure generates a mapping from a perceived ARMA process for prices to an actual process. A limited information rational expectations equilibrium is a fixed point of this mapping. We now fill in some technical details involved in constructing the mapping.

Define the *state* of the system, z_t, and the system noise ε_t as

$$z_t = \begin{bmatrix} y_t \\ w_t \\ x_t \\ u_t \end{bmatrix} \qquad \varepsilon_t = \begin{bmatrix} u_t \\ v_t \end{bmatrix}$$

Notice that both w_t and u_t are included in the state, where w_t is the innovation in the perceived law of motion. When the *perceived* law of motion for y_t is given by (6.3) and (6.4), then the *actual* law of motion for z_t can be computed to be

$$z_t = T(\beta) z_{t-1} + V(\beta)\varepsilon_t \tag{6.5}$$

where $\beta = [a\ c]$,

$$T(\beta) = \begin{bmatrix} -\lambda\Delta ca & -\lambda\Delta c^2 & \rho\Delta & d\Delta \\ -\lambda\Delta ca - a & -\lambda\Delta c^2 - c & \rho\Delta & d\Delta \\ 0 & 0 & \rho & d \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$V(\beta) = \begin{bmatrix} \Delta & \Delta \\ \Delta & \Delta \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \tag{6.6}$$

where $\Delta = (1 - \lambda a - \lambda c)^{-1}$.

Representation (6.5)–(6.6) gives the mapping from the parameters of the perceived law of motion in (6.3) for y to the actual law of motion for the entire state vector z_t. When (6.5) is the actual law of motion for z_t, for fixed β we can compute the covariance matrix of z_t associated with the stationary distribution of z_t. In particular, let

$$\Omega = E \begin{bmatrix} u_t \\ v_t \end{bmatrix} \begin{bmatrix} u_t \\ v_t \end{bmatrix}'$$

Let $M_z(\beta)$ be the covariance matrix $E z_t z_t'$ associated with the stationary distribution implied by (6.5) for fixed β. Then $M_z(\beta)$ satisfies the discrete Lyapunov equation

$$M_z(\beta) = T(\beta) M_z(\beta) T(\beta)' + V(\beta)\Omega V(\beta)' \tag{6.7}$$

The Lyapunov equation (6.7) can be solved for $M_z(\beta)$ using any of several algorithms.[8]

Consider the subset of z_t

$$z_{at} = \begin{bmatrix} y_t \\ w_t \end{bmatrix}$$

Denote the second moment matrix of z_{at} by $M_{z_a}(\beta)$. Evidently, $M_{z_a}(\beta)$ consists of the 2×2 submatrix in the upper left corner of $M_z(\beta)$. The covariance matrix $M_{zz_a}(\beta) = E z_t z_{at}'$ is the 4×2 submatrix consisting of the two leftmost columns of $M_z(\beta)$.

Notice that z_{at} is linked to z_t by

$$z_{at} = e_a z_t$$

where

$$e_a = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

We are interested in computing the projection of z_{at+1} on z_{at} when the law of motion for z_t is (6.5). Direct calculations establish that

$$E z_{at+1} | z_{at} = S(\beta) z_{at} \tag{6.8}$$

where

$$S(\beta) = e_a T(\beta) M_{zz_a}(\beta) M_{z_a}(\beta)^{-1} \tag{6.9}$$

The first row of $S(\beta)$ gives the coefficients in the projection of y_{t+1} on y_t and w_t, while the second row gives the coefficients in the projection of w_{t+1} on y_t and w_t. Thus, when the perceived projection of y_{t+1} on y_t and w_t is determined by parameters β, the actual projection of y_{t+1} on y_t and w_t is determined by the parameters $S_1(\beta)$, where $S_1(\beta)$ is the first row of $S(\beta)$.

2 Existence and Uniqueness of Stationary Equilibria

In this section[9] we describe the relationship between the fixed points of S_1 and limited information rational expectations equilibria. We state conditions for existence and uniqueness of a stationary limited information rational expectations equilibrium. We will show that stationary equilibria do not exist for some parameter values.

Previous papers analyzing convergence of least squares learning mechanisms using the ordinary differential equation approach have used two facts – (i) fixed points of the mapping S are the only possible limit points of a least squares learning mechanism and (ii) all fixed points of S correspond to rational expectations equilibria – both to establish that only rational expectations equilibria can be the limit points of the learning mechanism and to find conditions for convergence.[10] But in the model of this chapter, it can happen that fixed points of S_1 are not rational expectations equilibria.

Definition A stationary limited information rational expectations equilibrium (LIREE) is a fixed point of the mapping S_1, namely a pair of parameter values that satisfies $(a_f, c_f) = S_1(a_f, c_f)$ satisfying the following two conditions:

(a) the processes (y_t, w_t) generated by these parameters are measurable with respect to $(x_t, v_t, x_{t-1}, v_{t-1}, \ldots)$;

(b) w_t is measurable with respect to $(y_t, y_{t-1}, \ldots)$.

What is different from previous papers is the measurability requirements. These conditions are used because, as we will see in proposition 1, there exist fixed points of S_1 with a non-invertible w_t. The definition of S_1 above does not impose the natural requirement in rational expectations models that the prediction implied by the parameters (a, c) should be measurable with respect to past information and past exogenous variables.

Let us see in more detail the evolution of y_t and w_t at the fixed point. First of all, from equation (6.5) we have

$$y_t = \frac{1}{1-\lambda a}\left(\frac{1+dL}{1-\rho L}u_t + \lambda c w_t + v_t\right) \tag{6.10}$$

and

$$(1-\rho L)y_t = \frac{1}{1-\lambda a}\left[(1+dL)u_t + (1-\rho L)\lambda c w_t + (1-\rho L)v_t\right] \tag{6.11}$$

Notice that if we can find a white noise with finite variance w_t and a parameter $\bar{c}$ that satisfy

$$(1+\bar{c}L)w_t = \frac{1}{1-\lambda a}\left[(1+dL)u_t + (1-\rho L)\lambda \bar{c} w_t + (1-\rho L)v_t\right] \tag{6.12}$$

then combining (6.11) and (6.12) we know that y_t has an ARMA(1, 1) representation with white noise w_t.

Lemma 1 Let $\bar{c}$ be such that w_t satisfying (6.12) is a white noise with constant variance. Then $(a, c) = (\rho, \bar{c})$ is a fixed point of S_1.

Proof From equation (6.8), it is enough to check that the processes for y_t and w_t generated by the parameters $(\rho, \bar{c})$ satisfy $E(y_{t+1} \mid y_t, w_t) = \rho y_t + \bar{c} w_t$. Now, if w_t satisfies (6.12), we can write

$$E\left[y_{t+1}|y_t, w_t\right] = \rho y_t + c w_t + E\left[w_{t+1}|y_t, w_t\right]$$

and it is enough to check that $E\left[w_{t+1}|y_t, w_t\right] = 0$. Since by the assumptions of the lemma w_t is a white noise, we have

$$\begin{aligned}
\operatorname{cov}(w_{t+1}, y_t) &= \operatorname{cov}(w_{t+1}, \rho y_{t-1} + c w_{t-1} + w_t)\\
&= \rho \operatorname{cov}(w_{t+1}, y_{t-1})\\
&= \rho^i \operatorname{cov}(w_{t+1}, y_{t-i})
\end{aligned}$$

Letting i go to infinity we have that $\operatorname{cov}(w_{t+1}, y_t) = 0$.

Now it is clear that all we have to do to find fixed points of S_1 is to find values $\bar{c}$ that satisfy (6.12); the following proposition tells us what those are.

Proposition 1 There exist two fixed points of the mapping S_1 given by

$$\begin{aligned}
a_f &= \rho \qquad c_f = \frac{(1-\lambda\rho)\left[\delta - (\delta^2-1)^{1/2}\right]}{1+\lambda\left[\delta - (\delta^2-1)^{1/2}\right]}\\
\bar{a}_f &= \rho \qquad \bar{c}_f = \frac{(1-\lambda\rho)\left[\delta + (\delta^2-1)^{1/2}\right]}{1+\lambda\left[\delta + (\delta^2-1)^{1/2}\right]}
\end{aligned} \tag{6.13}$$

where

$$\delta = \frac{(1+d^2)\sigma_u^2 + (1+\rho^2)\sigma_v^2}{2\left(d\sigma_u^2 - \rho\sigma_v^2\right)}$$

Proof Equation (6.12) implies

$$[1-\lambda(\rho+c) + cL]w_t = [(1+dL)u_t + (1-\rho L)v_t] \tag{6.14}$$

This is the equation generating the w_ts in terms of the fundamentals. We want to find values of c that are consistent with w_t in (6.14) being a white noise. Taking the variance and the first autocovariance of both sides of (6.14) and using $\operatorname{cov}(w_t, w_{t-1}) = 0$, we have

$$\sigma_w^2\left\{[1-\lambda(\rho+c)]^2 + c^2\right\} = (1+d^2)\sigma_u^2 + (1+\rho^2)\sigma_v^2 \tag{6.15}$$

and

$$\sigma_w^2\left[(1-\lambda(c+\rho)\right]c = d\sigma_u^2 - \rho\sigma_v^2 \tag{6.16}$$

If $d\sigma_u^2 - \rho\sigma_v^2 = 0$ we see from equation (6.16) that the two solutions are given by $1-\lambda(c+\rho) = 0$ and $c = 0$. Otherwise, using (6.16), we can eliminate σ_w^2 from (6.15) and get

$$\left[\frac{c}{1-\lambda(\rho+c)}\right]^2 - 2\delta\frac{c}{1-\lambda(\rho+c)} + 1 = 0 \tag{6.17}$$

where δ has been defined in the statement of the theorem. This is a polynomial in $c/[1-\lambda(\rho+c)]$ with solutions given by

$$\frac{-c}{1-\lambda(p+c)} = -\delta \pm (\delta^2-1)^{1/2} \tag{6.18}$$

The formulas in the statement of the proposition follow immediately from (6.18).

To check that we have real solutions it is enough to check that $|\delta| > 1$. For convenience, let α be the right-hand side of (6.15) and β be the right-hand side of (6.16), so that $\delta = \alpha/2\beta$; then we have to check that $\alpha > 2|\beta|$. if $\beta > 0$ then

$$\alpha - 2\,|\,\beta\,| = (1-d)^2\sigma_u^2 + (1+\rho)^2\sigma_v^2 > 0 \tag{6.19a}$$

and if $\beta < 0$ then

$$\alpha - 2|\beta| = (1+d)^2\sigma_u^2 + (1-\rho)^2\sigma_v^2 > 0 \tag{6.19b}$$

so the solutions are real.

Proposition 1 gives the values of c that are consistent with (6.12). The value of w_t consistent with each c_f is found by performing forward or backward recursion on (6.12) depending on whether $-c/[1-\lambda(\rho+c)]$ is larger or smaller than unity in absolute value. So, using (6.14), we can set

$$\omega_t = \sum_{i=0}^{\infty}\left[\frac{-c}{1-\lambda(\rho+c)}\right]^i \frac{(1+dL)u_{t-i} + (1-\rho L)v_{t-i}}{1-\lambda(\rho+c)} \tag{6.20}$$

if $-c/[1-\lambda(\rho+c)]$ is less than unity in absolute value and

$$w_t = \sum_{i=1}^{\infty}\left[\frac{-c}{1-\lambda(\rho+c)}\right]^i \left[(1+dL)u_{t+i} + (1-\rho L)v_{t+i}\right] \tag{6.21}$$

otherwise.

This gives us the value of c_f and the corresponding innovation of y_t. To prove that w_t given by (6.20) or (6.21) is a white noise, simply observe that these satisfy (6.15) by construction, which holds if and only if $\mathrm{cov}(w_t, w_{t-1}) = 0$; that covariances with longer lags are zero follows immediately from (6.14).

Now the issue is which of these fixed points is a rational expectations equilibrium. First of all, it is clear that if w_t has to be written in terms of current and past ys then we will need that $|c| < 1$, but this may not be enough; the equation that gives us the evolution of the ws in terms of the fundamentals is (6.14). Clearly, if $-c/[1-\lambda(\rho+c)]$ is larger than unity in absolute value w_t will not be measurable with respect to past exogenous variables. These requirements are formalized in the following.

Proposition 2 Each fixed point of proposition 1 is an LIREE if and only if $|\,c\,| < 1$ and

$$\left|\frac{-c}{1-\lambda(\rho+c)}\right| < 1$$

Proof We can write w_t in terms of past ys by setting $w_t = \sum_{i=0}^{\infty} c^i(1-\rho L)y_{t-i}$, but this sum is convergent if and only if $|\,c\,| < 1$. Similarly, if $-c/[1-\lambda(\rho+c)]$ is larger than unity in absolute value equation (6.21) tells us that w_t will depend on future values of the exogenous variables.

Finally, we come to the characterization of the LIREEs in terms of these fixed points of S_1. The next proposition says that (ρ, c_f) (i.e. the first fixed point in proposition 1) is the only candidate for being an LIREE because the fixed point we have labeled (ρ, c_f) does not satisfy the conditions of proposition 2. Also, this proposition says that if (ρ, c_f) does not satisfy the conditions of proposition 2 then there is no equilibrium.

Proposition 3

(a) If $\delta > 0$ an LIREE exists if and only if c_f satisfies the conditions of proposition 2 ($\bar{c}_f$ if $\delta < 0$).

(b) When an LIREE exists, the processes y_t, w_t generated by $(a, c) = (\rho, c_f)$ are the unique rational expectations equilibrium with limited information.

Proof We first prove part (a) for $\delta > 0$. The statement that if c_f satisfies the conditions of proposition 2 then an LIREE exists follows immediately from proposition 2.

Now we observe that

$$\left|\frac{-\bar{c}_f}{1-\lambda(\rho+\bar{c}_f)}\right| = |\delta + (\delta^2-1)^{1/2}| > |\delta| > 1$$

where the last inequality has been proved in proposition 1. Therefore, $\bar{c}_f$ cannot be an LIREE, and the only fixed point that can satisfy all the conditions for an LIREE is c_f. If c_f does not satisfy the inequalities in proposition 2, then no stationary equilibrium exists. A stationary equlibrium would have to satisfy equations (6.12) and (6.14), and only (ρ, c_f) and $(\rho, \bar{c}_f)$ satisfy these equations, but we have just ruled out $\bar{c}_f$ as an equilibrium. This argument also proves part (b), because if c_f satisfies the inequalities of proposition 2 the argument in the previous paragraph proves that there exists no other equilibrium.

It is possible to find parameter values for which no stationary equilibrium exists.

This is not surprising in view of the work of Futia (1981), who studied a version of our model in which $\sigma_v^2 = 0$. For some parameter values Futia found that no stationary equilibrium exists. For our ARMA process for x_t, and if $\sigma_v^2 = 0$, the value for c at equilibrium is

$$\frac{d(1-\lambda\rho)}{1+\lambda d} \tag{6.22}$$

For some values of the parameters this can be larger than unity (e.g. if $d = 0.5$, $\lambda = -0.9$, $\rho = 0.8$), and it is easy to show that for σ_v^2 small the value of c_f gets arbitrarily close to that given by (6.22).

Nevertheless, particular conditions on the parameters of the model can be imposed to guarantee that there exists a unique stationary equilibrium. Some of these conditions are the following.

Proposition 4 Each of the following set of conditions is sufficient for existence of a unique stationary LIREE:

(a) $\lambda > 0$

(b) $d = 0$

(c) σ_v^2 arbitrarily large

(d) σ_v^2 arbitrarily small and the expression in (6.22) is less than unity in absolute value.

Proof The proofs involve simple algebra and most of the details will be omitted. We only give details for the case $\delta > 0$.

(a) We first need to show that $\delta - (\delta^2 - 1)^{1/2}$ is less than unity in absolute value. This follows from the fact that this is an increasing function of δ, the fact that $|\delta| > 1$ (which has been shown in the proof of proposition 1) and the fact that $-1 < -\delta + (\delta^2 - 1)^{1/2} < 0$. Now, since $\lambda > 0$, $|\rho| < 1$ and $\delta - (\delta^2 - 1)^{1/2} > 0$, we have

$$0 < c_f < \frac{1-\lambda\rho}{1+\lambda} < 1$$

and both conditions of proposition 2 are satisfied.

(c) Take $\rho < 0$. As σ_v^2 goes to infinite δ goes to $-(1+\rho^2)/\rho$, $\delta - (\delta^2 - 1)^{1/2}$ goes to $-\rho$ and c_f goes to $-\rho$, which is less than unity in absolute value.

This characterizes in some detail the stationary equilibria, the fixed points of S_1 and their relationship. There could be more fixed points of S_1 but they could not be LIREE because they do not satisfy the requirements in proposition 2. Also, there might be rational expectations equilibria involving more error terms. Finally, we note that, when one exists, the LIREE studied in this section is of full order, in the sense used by Sargent (1991).

3 Learning

We now turn to a learning version of the model. We continue to define $T(\beta)$ and $V(\beta)$ as in (6.6). The law of motion of z_t is now given by

$$z_t = T(\beta_{t-1})z_{t-1} + V(\beta_{t-1})\varepsilon_t \tag{6.23}$$

where $\beta_t = (a_t, c_t)$. The parameters (a_t, c_t) are estimators of (a, c) in (6.4). Agents behave as though they live in a time-invariant system, though (6.23) belies that belief. The parameters are estimated via one of the recursive algorithms described by Ljung and Söderström (1983). We consider two possible estimators: (i) pseudo-linear regression, and (ii) the recursive prediction error method.

Pseudo-linear regression

Under pseudo-linear regression, the system evolves according to

$$\hat{y}_t = a_{t-1}y_{t-1} + c_{t-1}\hat{w}_{t-1} \tag{6.24a}$$

$$\psi_t = \begin{bmatrix} y_{t-1} \\ \hat{w}_{t-1} \end{bmatrix} \tag{6.24b}$$

$$\hat{w}_t = y_t - \hat{y}_t \tag{6.24c}$$

$$\gamma_t = 1/t \tag{6.24d}$$

$$R_t = R_{t-1} + \gamma_t \left[\psi_t \psi_t' - R_{t-1}\right] \tag{6.24e}$$

$$\begin{bmatrix} a_t \\ c_t \end{bmatrix} = \begin{bmatrix} a_{t-1} \\ c_{t-1} \end{bmatrix} + \gamma_t R_t^{-1} \psi_t \hat{w}_t \tag{6.24f}$$

$$y_t = e_1 z_t \tag{6.24g}$$

$$\beta_{t-1} = (a_{t-1}\ c_{t-1}) \tag{6.24h}$$

$$z_t = T(\beta_{t-1})z_{t-1} + V(\beta_{t-1})\varepsilon_t \tag{6.24i}$$

Recursive prediction error method

The system is identical to that under the pseudo-linear regression except that the second equation in the system, (6.24), is altered to

$$\psi_t = -c_{t-1}\psi_{t-1} + \begin{bmatrix} y_{t-1} \\ \hat{w}_{t-1} \end{bmatrix} \tag{6.24b$'$}$$

For estimating the parameters of an *exogenous* ARMA(1, 1) process, the recursive prediction error method has an interpretation as a recursive optimal instrumental variable estimator. Both pseudo-linear regression and the recursive prediction error method are devices for recursively estimating parameters via stochastic approximation on the orthogonality condition $Ew_t\psi_t = 0$. Pseudo-linear regression chooses ψ_t to impose that w_t be orthogonal only to

$$\begin{bmatrix} y_{t-1} \\ w_{t-1} \end{bmatrix}$$

while the recursive prediction error method forms the instrument ψ_t as the geometric distributed lag of

$$\begin{bmatrix} y_{t-1} \\ w_{t-1} \end{bmatrix}$$

given by (6.24b′). It can be shown that ψ_t given by (6.24b′) is an optimal form of instrument for an ARMA(1, 1) model.[11]

The associated ordinary differential equations

Application of the apparatus of Marcet and Sargent (1989a, b) can be used to find systems of ordinary differential equations whose limiting behavior governs the limiting behavior of the systems of stochastic difference equations (6.24). For each algorithm, there is a "large" ordinary differential equation that governs the global convergence of a version of the algorithm which has been altered by the addition of a "projection facility" that instructs the algorithm to ignore observations that threaten to drive β_t, R_t outside of a prescribed set. There is also a "small" ordinary differential equation whose behavior governs the limiting behavior of β_t in the locality of a fixed point. We provide a formal statement of convergence theorems in the appendix. These theorems are simple adaptations of propositions in Marcet and Sargent (1989a, b) to the current environment. In the remainder of this chapter, we shall describe these associated ordinary differential equations.

Pseudo-linear regression

For pseudo-linear regression, the ordinary differential equation system is

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\beta' &= R^{-1}E\psi_t\hat{w}_t \\ \frac{\mathrm{d}}{\mathrm{d}t}R &= M_{z_a}(\beta) - R \end{aligned} \tag{6.25}$$

where $M_{z_a}(\beta) = Ez_{at}z_{at}'$. For $\psi_t = z_{at-1}$, as under pseudo-linear regression, the first equation of (6.25) can be rewritten as

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\beta' &= R^{-1}Ez_{at-1}\left[e_1T(\beta)z_{t-1} + e_1V(\beta)\varepsilon_t - \beta z_{t-1}^a\right] \\ &= R^{-1}Ez_{at-1}\left[z_{t-1}'T(\beta)'e_1 - z_{at-1}'\beta'\right] \\ &= R^{-1}M_{z_a}(\beta)\left[M_{z_a}(\beta)^{-1}M_{z_a,z}(\beta)T(\beta)'e_1' - \beta'\right] \end{aligned}$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t}\beta' = R^{-1}M_{z_a}(\beta)[S(\beta)'u_1' - \beta'] \tag{6.26}$$

where

$$S(\beta) = e_aT(\beta)M_{z,z_a}(\beta)M_{z_a}(\beta)^{-1} \tag{6.27}$$

In (6.26), u_1 selects the first row of $S(\beta)$.

In summary, under pseudo-linear regression, we have the ordinary differential equation

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\beta' &= R^{-1}M_{z_a}(\beta)\left[S(\beta)'u_1' - \beta'\right] \\ \frac{\mathrm{d}}{\mathrm{d}t}R &= M_{z_a}(\beta) - R \end{aligned} \tag{6.28}$$

Recursive prediction error method

Under the recursive prediction error method, the ordinary differential equation is

$$\frac{\mathrm{d}}{\mathrm{d}t}\beta' = R^{-1}E\psi_t\hat{w}_t$$

$$\frac{\mathrm{d}}{\mathrm{d}t}R' = E\psi_t\psi_t'(\beta) - R$$

The first equation can be represented as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\beta' &= R^{-1}E\psi_t(y_t - \beta z_{at-1})\\ &= R^{-1}E\psi_t\left[e_aT(\beta)z_{t-1}' + e_1V(\beta)\varepsilon_t - \beta z_{at-1}\right]\\ &= R^{-1}E\psi_t\left[z_{t-1}'T(\beta)'e_a' - z_{at-1}'\beta'\right]\\ &= R^{-1}\left[E\psi_t z_{t-1}'T(\beta)'e_a' - E\psi_t z_{at-1}'\beta'\right]\\ &= R^{-1}E\psi_t z_{at-1}'\left[(E\psi_t z_{at-1}')^{-1}(E\psi_t z_{at-1}')T(\beta)'e_a' - \beta'\right]\end{aligned}$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t}\beta' = R^{-1}E\psi_t z_{at-1}'\left[P(\beta)'u_1' - \beta'\right]$$

where

$$P(\beta) = e_aT(\beta)M_{z,\psi}(\beta)M_{z_a,\psi}(\beta)^{-1} \tag{6.29}$$

In summary, under the recursive prediction error method, the ordinary differential equation is[12]

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\beta' &= R^{-1}M_{\psi,z_a}(\beta)\left[P(\beta)'u_1' - \beta'\right]\\ \frac{\mathrm{d}}{\mathrm{d}t}R &= M_{\psi}(\beta) - R\end{aligned} \tag{6.30}$$

The ordinary differential equations (6.28) and (6.30) play the roles of the "large ordinary differential equations" in the analysis of Marcet and Sargent (1989a, b). If we can find a set in the space in which (β_t, R_t) lives such that the large ordinary differential equation has a unique fixed point and is globally stable within that set, then we can find a modified version of our recursive algorithms (6.24) that

converges strongly to that fixed point. The modification of the algorithm involves imposing a "projection facility" that instructs the algorithm to ignore observations that threaten to drive the parameters outside the set just described. The appendix contains formal statements of the convergence results that can be attained for our systems.

The operators P and S share a fixed point

The operators P and S associated with the recursive prediction error method and pseudo-linear regression, respectively, share a fixed point. We formulate this fact in terms of the following proposition.

Proposition 5 Suppose that β_f satisfies $\beta_f = S_1(\beta_f)$. Then $\beta_f = P_1(\beta_f)$.

Proof We have noted that $\beta_f = S_1(\beta_f)$ implies that $w_t(\beta_f)$ is an innovation for y_t relative to y^{t-1}. This implies that $E\psi_{t-1}(\beta_f)w_t = 0$ which implies that $\beta_f = P_1(\beta_f)$.

Interpretation of $S_1(\beta)$ and $P_1(\beta)$

Consider the regressions

$$z_t = \gamma z_{at} + r_t \qquad Ez_{at}r_t' = 0$$

$$z_t = \phi z_{at} + \tilde{r}_t \qquad E\psi_t\tilde{r}_t' = 0$$

The normal equations for these two regressions are

$$\begin{aligned}\gamma &= M_{z,z_a}(\beta)M_{z_a}(\beta)^{-1}\\ \phi &= M_{z,\psi}(\beta)M_{z_a,\psi}(\beta)^{-1}\end{aligned} \tag{6.31}$$

Here γ is the "ordinary least squares" estimator of the regression equation, while ϕ is an "instrumental variables" estimator.

Notice that we can represent $S_1(\beta)$ and $P_1(\beta)$ as follows:

$$\begin{aligned}S_1(\beta) &= u_1T(\beta)\gamma\\ P_1(\beta) &= u_1T(\beta)\phi\end{aligned} \tag{6.32}$$

Formulas for moment matrices

To compute $P_1(\beta)$, we need formulas for $M_{z,\psi}(\beta) = Ez_{t-1}\psi_t'$ and $M_{z_a,\psi}(\beta) = Ez_{at-1}\psi_t'$. To obtain these, we first use (6.5) to compute

$$Ez_{t+j}z_{at}' = T(\beta)^j M_z(\beta)e_a' \tag{6.33}$$

Next, we have from the definition of ψ_t in the pseudo-linear regression (henceforth denoted RPEM) (equation (6.24b')) that

$$\begin{aligned} Ez_{t-1}\psi_t' &= \sum_{j=0}^{\infty}(-c)^j Ez_{t-1}z_{at-j-1}' \\ &= \sum_{j=0}^{\infty}[-cT(\beta)]^j M_z(\beta)e_a' \end{aligned} \tag{6.34}$$

or

$$Ez_{t-1}\psi_t' = [I + cT(\beta)]^{-1}M_z(\beta)e_a' \tag{6.35}$$

We also have

$$Ez_{at-1}\psi_t' = e_a[I + cT(\beta)]^{-1}M_z(\beta)e_a' \tag{6.36}$$

Formulas for $S_1(\beta)$ and $P_1(\beta)$

Using the above formulas for the moment matrices, we have the following formulas for $P_1(\beta)$ and $S_1(\beta)$:

$$\begin{aligned} P_1(\beta) &= e_1T(\beta)\{[I + cT(\beta)]^{-1}M_z(\beta)e_a'\} \\ &\quad \times \{e_a[I + cT(\beta)]^{-1}M_z(\beta)e_a'\}^{-1} \end{aligned} \tag{6.37}$$

$$S_1(\beta) = e_1T(\beta)[M_z(\beta)e_a'][e_aM_z(\beta)e_a']^{-1} \tag{6.38}$$

Local analysis of the ordinary differential equations

Recursive prediction error method

Consider the "large" ordinary differential equation for the RPEM:

$$\frac{\mathrm{d}}{\mathrm{d}t}\beta' = R^{-1}M_{\psi,z_a}(\beta)[P(\beta)'e_1' - \beta']$$

$$\frac{\mathrm{d}}{\mathrm{d}t}R = M_\psi(\beta) - R$$

In the vicinity of a fixed point β_f of $P_1(\beta_f)$, this system has dynamics that are governed by a version of proposition 3 of Marcet and Sargent (1989a). In particular, we have to study the matrix

$$\mathcal{M}_P = \frac{\partial \text{ col}}{\partial \text{ col } \beta'}\left\{R^{-1}M_{\psi,z_a}(\beta)'\left[P_1(\beta)' - \beta'\right]\right\}\Big|_{\beta=\beta_f}$$

Computing the indicated derivative and evaluating at $\beta = \beta_f$ gives

$$\begin{aligned} \mathcal{M}_P = R^{-1}&\frac{\partial \text{ col}}{\partial \text{ col } \beta'}M_{\psi,z_a}(\beta)'\left[P_1(\beta_f)' - \beta_f'\right] \\ &+ R^{-1}M_{\psi,z_a}(\beta_f)'\left[\frac{\partial \text{ col } P_1(\beta)'}{\partial \text{ col } \beta'} - I\right]\Big|_{\beta_f} \end{aligned}$$

or

$$\mathcal{M}_P = R^{-1}M_{\psi,z_a}(\beta)'\left\{\left[\frac{\partial \text{ col } P_1(\beta)'}{\partial \text{ col } \beta'}\right]_{\beta=\beta_f} - I\right\}$$

because $P_1(\beta_f) = \beta_f$.

To check the local stability of the RPEM, we have to compute $\mathcal{M}_P$ and check whether its eigenvalues are all strictly negative in real part.

Pseudo-linear regression

Consider the "large" ordinary differential equations (6.28) for pseudo-linear regression:

$$\frac{\mathrm{d}}{\mathrm{d}t}\beta' = R^{-1}M_{z_a}(\beta)\left[S(\beta)'u_1' - \beta'\right]$$

$$\frac{\mathrm{d}}{\mathrm{d}t}R = M_{z_a}(\beta) - R$$

The dynamics of the algorithm in the vicinity of β_f are governed by

$$\mathcal{M}_S = \frac{\partial \text{ col}}{\partial \text{ col}\,\beta'} \left\{ R^{-1} M_{z_a}(\beta)' \left[S_1(\beta)' - \beta' \right] \right\} \Big|_{\beta=\beta_f}$$

We can compute

$$\mathcal{M}_S = \left[\frac{\partial \text{ col } S_1(\beta)'}{\partial \text{ col } \beta'} \right]_{\beta=\beta_f} - I$$

because, at $\beta = \beta_f$, $R^{-1} M_{z_a} = I$. To determine the local stability of the system under pseudo-linear regression, we can compute $\mathcal{M}_S$ and check whether its eigenvalues are all strictly negative in real part.

Simulations

In this section, we describe solutions of the large ordinary differential equation (6.30) for the RPEM for two sets of parameter values. We also report a simulation of the system operating under the RPEM. For our first parameter set, we choose $\lambda = 0.75$, $\rho = 0.8$, $d = -0.95$, $\sigma_u^2 = \sigma_v^2 = 1$, $\sigma_{uv} = 0$. For these parameter values, the equilibrium values are $a = 0.8$, $c = -0.9559$, and for the recursive prediction error method

$$R = M_\psi = \begin{bmatrix} 4.5530 & 6.9665 \\ 6.9665 & 19.0026 \end{bmatrix}$$

For these parameter values, we calculated that the eigenvalues of $\mathcal{M}_P$ at the fixed point are $(-0.3924, -0.1035)$ and that the eigenvalues of $\mathcal{M}_S$ are $(-0.4002 \pm 0.4517i)$. For starting values of $a(0) = 0.1$, $c(0) = 0$, $R(1,1) = 20$, $R(2,2) = 30$, $R(1,2) = 20$, we solved the large ordinary differential equation (6.30) for the RPEM by using a Runge–Kutta algorithm.[13] Figures 6.1 and 6.2 plot the solution. Evidently, (β, R) is converging to equilibrium values.

Figures 6.3–6.5 describe the solutions of (6.30) for a second parameter set which is equal to the first except that now we set $d = 0$. Here the equilibrium values are $a = 0.8$, $c = -0.1808$, and for the RPEM

$$R = M_\psi = \begin{bmatrix} 22.9489 & 9.6586 \\ 9.6586 & 8.5408 \end{bmatrix}$$

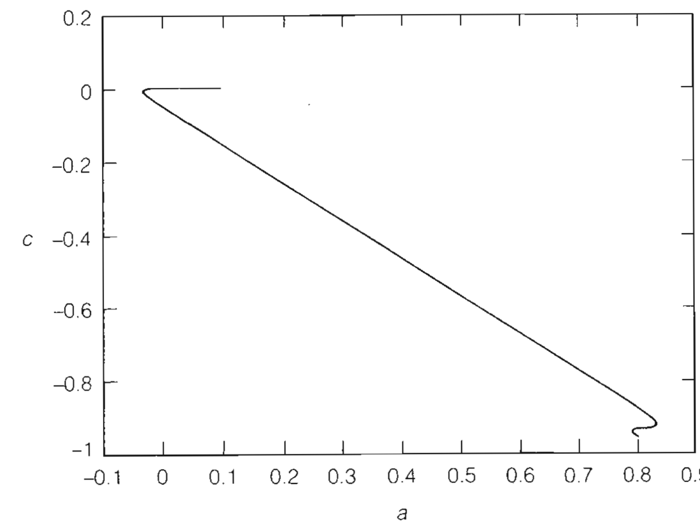

Figure 6.1 Parameter $d = -0.95$. Plot of a versus c determined by the big ordinary differential equation for the recursive prediction error method

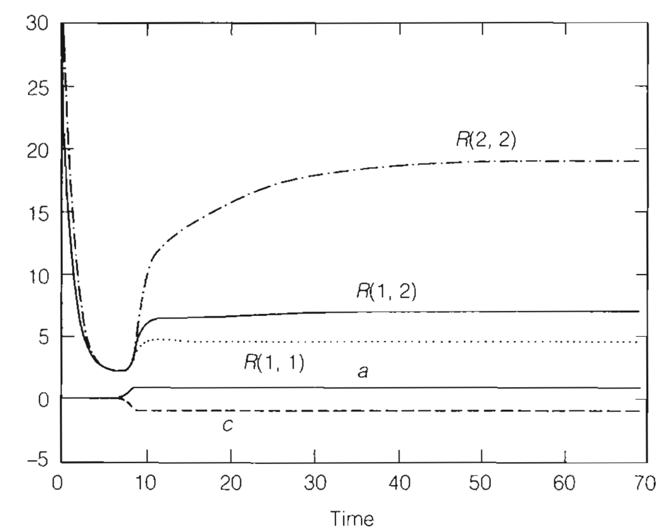

Figure 6.2 Parameter $d = -0.95$. Plot of a, c, amd M_ψ as determined by the big ordinary differential equation for the recursive prediction error method

For these parameters, we calculated that the eigenvalues of $\mathcal{M}_S$ are $(-1.1017 \pm 0.2084i)$, while those for $\mathcal{M}_P$ are $(-1.0922 \pm 0.4835i)$. Figures 6.3–6.5 give the solutions of (6.30) for the same initial conditions used above. The convergence to equilibrium is more rapid now, which is consistent with the smaller eigenvalues of $\mathcal{M}_P$ for the second set of parameter values.

Figures 6.6 and 6.7 report the results of simulating the recursive prediction error method for the above parameter settings using a pseudo random number generator to produce Gaussian εs. For this simulation, we set initial conditions of $a(0) = 0.1$, $c(0) = 0$,

$$R = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$$

We set the initial value of t in the simulation at $t = 100$, and simulated the system out to $t = 5000$.[14] The simulated paths for a and c (and also those for R, which are not shown) seem to be converging to their equilibrium values. Notice how qualitatively figures 6.6 and 6.7 resemble figures 6.3 and 6.4, respectively.

We have computed solutions of the differential equation for many other parameter values and initial conditions. For all the values that we have checked, the eigenvalues of the $\mathcal{M}$s were always negative in real part, and the solutions of the big ordinary differential equation always converged to the equilibrium for

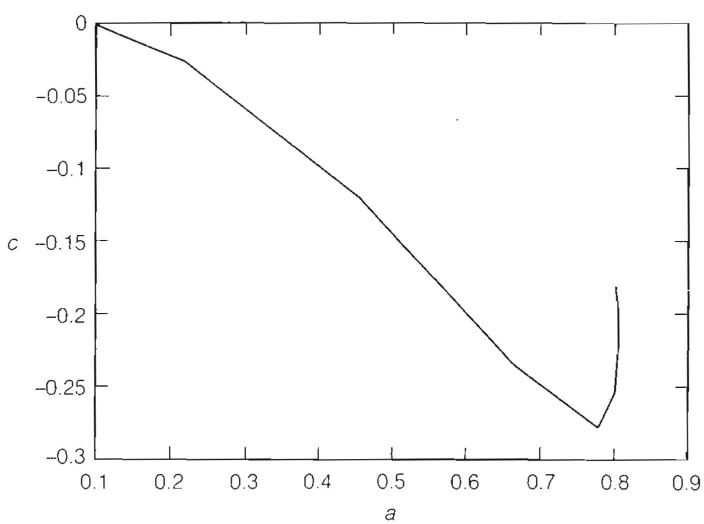

Figure 6.3 Parameter $d = 0$. Plot of a versus c determined by the big ordinary differential equation for the recursive prediction error method

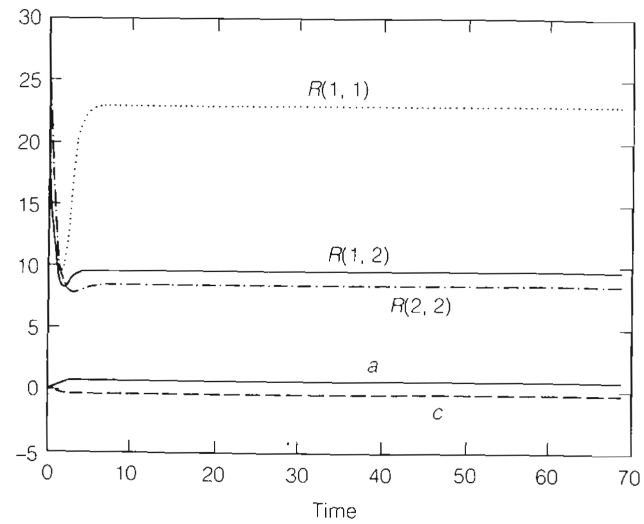

Figure 6.4 Parameter $d = 0$. Plot of a, c, and M_ψ as determined by the big ordinary differential equation for the recursive prediction error method

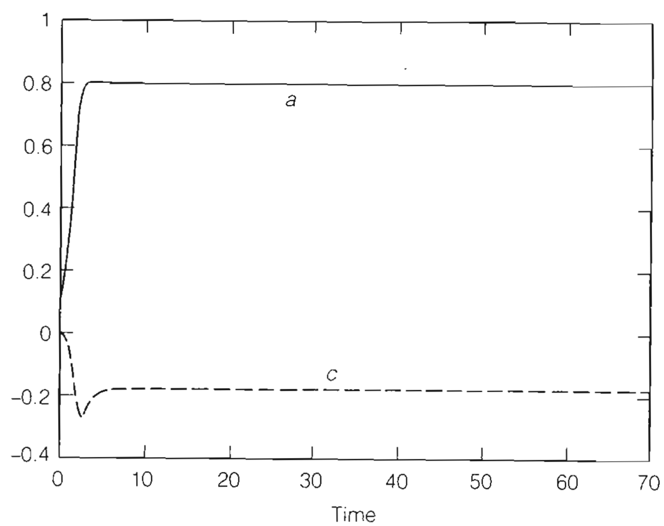

Figure 6.5 Parameter $d = 0$. Plot of a and c determined by the big ordinary differential equation for the recursive prediction error method

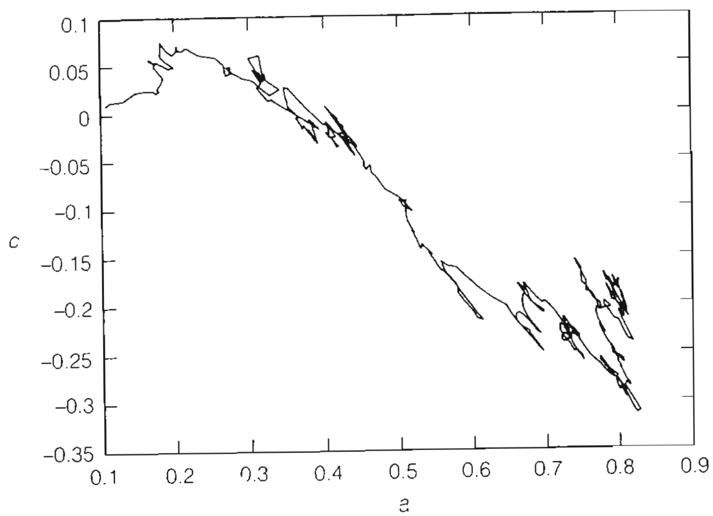

Figure 6.6 Parameter $d = 0$. Plot of a versus c for a simulation of the system with the recursive prediction error method

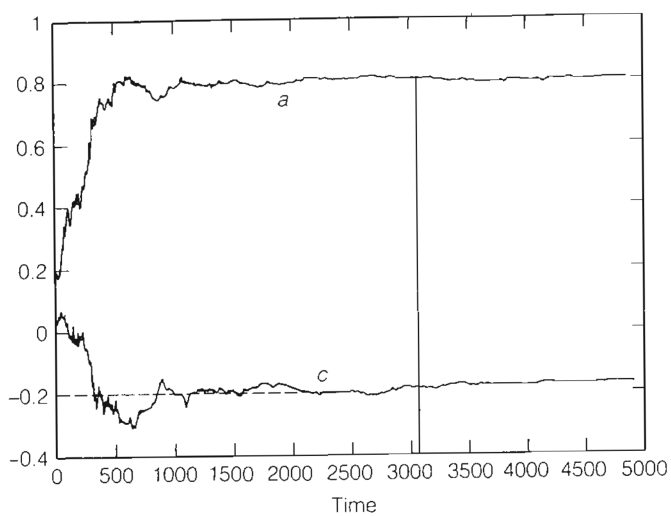

Figure 6.7 Parameter $d = 0$. Plot of a and c for a simulation of the system with the recursive prediction error method

alternative starting values that satisfied $a < 1$, $c < 1$. We also computed solutions for the system governed by pseudo-linear regression, with qualitatively similar results. As with the above two settings of parameter values, the real parts of the eigenvalues of the relevant $\mathcal{M}$s indicated slightly slower rates of convergence for pseudo-linear regression.

The propositions stated in the appendix and in Marcet and Sargent (1989a, b) provide more details about the senses in which the limiting properties of our learning systems can be discovered by studying their associated ordinary differential equations. These propositions support the following conclusions about the model of this paper.

(i) A version of Margaret Bray's (1982) result holds, stating that the only possible limit point of one of our learning algorithms is an LIREE.

(ii) Global convergence of the algorithms to a rational expectations equilibrium depends on the behavior of the "big" ordinary differential equation at the boundary of the set D_1 defining the "projection facility." Almost sure convergence depends on the trajectories of the ordinary differential equation pointing toward the interior of the set D_1. Even for models as simple as ours, the big ordinary differential equation has a five-dimensional state vector, causing us to resort to numerical methods to check the behavior of the trajectories.

(iii) Local stability is governed by the eigenvalues associated with a smaller ordinary differential equation.

4 Speed of Convergence

In this section we describe some results on the rate of convergence that we attain by applying a new theorem by Benveniste, Métivier and Priouret. We also describe a numerical procedure for estimating the rate of convergence by simulations. We first apply this procedure to the model of section 1 maintaining the hidden information assumption. Then we consider a full information case.

Analytic results

During the last decade our understanding of what determines convergence of least squares learning schemes in a self-referential dynamic econonic model has increased considerably. Our knowledge about the speed of convergence, however, is very limited.[15]

A relatively new result in Benveniste et al. (1990) (theorem 3, page 110) seems to be the most powerful result up to date. Consider an on-line algorithm that obeys

$$\beta_t = \beta_{t-1} + \frac{1}{t} Q(\beta_{t-1}, z_t)$$

Let $h(\beta) = E[Q(\beta, z_t)]$, where z_t in this expectation represents the process that obeys equation (6.5) for β given, and let β be such that $h(\beta) = 0$. The theorem of Benveniste et al. concludes that if the derivative of $h(\beta)$ has all eigenvalues less than $-1/2$ in real part then

$$t^{0.5}(\beta_t - \beta_f) \xrightarrow{\mathcal{D}} \mathrm{N}(0, P)$$

where the matrix P satisfies

$$\left[\frac{I}{2} h_\beta(\beta_f)\right] P + P \left[\frac{I}{2} h_\beta(\beta_f)\right]' + EQ(\beta_f, z_t) Q(\beta, z_t)' = 0$$

Thus, if the above conditions are met, we have root-t convergence as in the classical statistics case, although the formula for the variance of the estimators β is modified due to the presence of the terms depending on h_β. Notice that in the classcial case h_β is equal to the identity and P is the classical variance–covariance matrix. Also, we see that, for higher eigenvalues of $h_\beta(\beta_f)$, convergence is slower in the sense that the asymptotic variance–covariance matrix of the limiting distribution is higher.

Applying these results to least squares learning, we know that $h_\beta(\beta_f) = \partial S_1(\beta_f)/\partial\beta - I$ (see Marcet and Sargent, 1989b, proposition 1, statement iv), so that the condition to apply the theorem by Benveniste et al. translates into all eigenvalues of $\partial S_1(\beta_f)/\partial\beta$ being less than 1/2 in real part, which delivers root-t convergence.[16]

When this condition on the eigenvalues of the derivative of S_1 is not met we know of no analytic results on asymptotic distributions that we can apply here. The reason the proof of the Benveniste et al. theorem does not apply is that the importance of initial conditions fails to die out at an exponential rate as is needed for root-t convergence. Intuitively, root-t convergence obtains if the autocovariance of a process is summable, which means that the effect of initial conditions evaporates at an exponential rate, and this only happens if h_β is low. All of this suggests that, if $h_\beta(\beta_f)$ is too large, $\beta_t - \beta_f$ may go to zero at a rate slower than root t, unlike the usual case with classical estimation in stationary time series processes.[17] In other words, if the derivative of S_1 is too large, we expect $t^{\bar\delta}(\beta_t - \beta_f)$ to go to zero only for $\bar\delta \leq \delta < 1/2$.

Rates of convergence by simulation

In this section we describe a numerical procedure for exploring the rate of convergence by simulation. The Monte Carlo calculations of the rate of convergence are based on the *assumption* that there is a δ for which

$$t^\delta(\beta_t - \beta_f) \xrightarrow{\mathcal{D}} F \tag{6.39}$$

for some nondegenerate well-defined distribution F with mean zero. Then $t^{\bar\delta}(\beta_t - \beta_f) \to 0$ for $\bar\delta < \delta$, and we will call δ the rate of convergence of $\{\beta_t\}$.

Letting σ_F^2 denote the variance under the distribution F, (6.39) implies that $E[t^\delta(\beta_t - \beta_f)]^2 \to \sigma_F^2$ as $t \to \infty$. Therefore,

$$\frac{E[t^\delta(\beta_t - \beta_f)]^2}{E[(kt)^\delta(\beta_{tk} - \beta_f)]^2} \to 1$$

which, in turn, implies that

$$\frac{E(\beta_t - \beta_f)^2}{E(\beta_{tk} - \beta_f)^2} \to k^{2\delta} \quad \text{as } t \to \infty$$

This justifies using

$$\delta = \frac{1}{\log k} \log \left[\frac{E(\beta_t - \beta_f)^2}{E(\beta_{tk} - \beta_f)^2}\right]^{1/2} \tag{6.40}$$

for large t as an approximation to the rate of convergence. Given t and k, the expectations involved can be approximated by Monte Carlo integration, i.e. by simulating a large number N of independent realizations of length t and tk and calculating the mean square error across realizations.

Rates of convergence with hidden and full information

We now analyze numerically the rate of convergence in the model of the previous sections with and without hidden information. We start analyzing the version of the model with hidden information and agents using ARMA learning schemes, as in section 1. It will turn out that, with this informational structure, this model is not sufficient to study the most interesting issue, because root-t convergence *always* seems to hold because the relevant eigenvalues are always less than $-1/2$ for this model. This prompts us to look at a version of the model with full information, where agents see the shocks u_t and v_t.

In all the simulations we calculated the rates of convergence with 1000 independent realizations. We used three different seeds of the random number generator, and the rates of convergence were within about 0.03 of each other. For both versions of the model the initial conditions for the parameters were set equal to the limiting point, so that $\beta_0 = \beta_f$; for the matrix R_0 we used the second moment matrix at the fixed point multiplied by a hundred, so this is the true proportions of the elements of this matrix with as much weight as if we had had 100 observations at this point; therefore $R_0 = M_z(\beta_f) \times 100$. We have also performed simulations with initial conditions away from the fixed point of S; the results on the *rate* of convergence are not affected by the choice of initial conditions, although they slow down convergence considerably, particularly in the cases with a large derivative of S, as in the model with full information. For the case of hidden information we used a projection facility that ignored observations that led the beliefs about a_t to be larger than $(a_f + 1)/2$.

Table 6.1 reports the rates of convergence for the model of section 1 with hidden information and the least squares learning scheme (pseudo linear regression).

Table 6.1 Hidden information and ARMA learning with pseudo-linear regression: $\rho = 0.9$; $\sigma_u = \sigma_v = 0.1$; $d = 0$

λ	δ t = 500 to 2000	δ t = 2000 to 10,000	Eigenvalues of S_β in real part
0.1	0.476	0.475	–0.051, –0.342
0.145	0.473	0.474	–0.082, –0.337
0.19	0.471	0.473	–0.158, –0.249
0.235	0.467	0.473	–0.195
0.28	0.464	0.473	–0.190
0.325	0.461	0.473	–0.185
0.37	0.457	0.473	–0.182
0.415	0.454	0.472	–0.176
0.46	0.450	0.472	–0.177
0.505	0.446	0.471	–0.177
0.55	0.443	0.471	–0.185
0.595	0.438	0.471	–0.192
0.64	0.435	0.470	–0.204
0.685	0.430	0.470	–0.227
0.73	0.426	0.470	–0.253
0.775	0.421	0.470	–0.291
0.82	0.417	0.470	–0.348
0.865	0.410	0.470	–0.440
0.91	0.403	0.470	–0.567
0.955	0.053	0.442	–0.314, –1.24

These rates are calculated with the Monte Carlo method described above. Each table uses parameter values $\rho = 0.9$, $\sigma_u = \sigma_v = 0.1$, $d = 0$, but λ varies in small increments. We report the eigenvalues in real part of the derivative of S_1 for each value of λ; they are all negative, so that the theorem of Benveniste et al. applies.[18] Our calculations show that the numerical rate of convergence is very close to 1/2 when the length of the observations goes from 2000 to 10,000, but the rate can be much smaller below 2000; in fact, the rate is smaller the larger is λ. So the assertion of the theorem of root-t convergence seems to be nearly true in samples of about 10,000. It is remarkable, though, that in samples of smaller size the rate of convergence can be very low; in particular, for the highest λ there is almost no improvement in mean square error when going from a length of 500 to 2000 observations.

Table 6.2 takes the same model and the same informational structure as the previous table, but it uses the learning scheme based on the recursive prediction error method. We see that the eigenvalues are even more negative than in the previous table, so that the Benvenite et al. theorem applies, and we have root-t convergence.

Since the eigenvalues of the derivative of S and P are always negative in tables 6.1 and 6.2, the rates of convergence there can be used to illustrate the short sample properties of the model, and to see if the asymptotic distribution is a good approximation. We calculated the eigenvalues of S and P for many different parameter settings of the model and we always found that the eigenvalues of the derivatives had negative real parts. This means that if we use the model of section 1, with hidden information and ARMA learning schemes, we cannot explore our conjecture of the previous section that, the larger the derivative at β_f, the slower is the rate of convergence when the theorem by Benvenite et al. does not apply. For this purpose we modify the model slightly and assume that agents observe all shocks dated t or earlier (including us and vs), so that they form expectations using the only relevant information, namely x_t, and their expectations about the future are given by $\hat{E}_t(y_{t+1}) = \beta_t x_t$, where β_t is the ordinary least squares estimate of a regression coefficient of y_{t+1} on x_t.[19] Then this becomes a minor complication in example d in section 4 of Marcet and Sargent (1989a), and it is easy to check that the mapping S (identical to the mapping T in this example) and the fixed point β_f are given by

$$S(\beta) = \rho(1 + \lambda\beta) \qquad \beta_f = \rho/(1 - \lambda\rho)$$

so that $\partial S(\beta_f)/\partial\beta = \rho\lambda$.

Table 6.3 reports calculations for the same parameter values as tables 6.1 and 6.2. We can see how the rate of convergence is very slow for high values of λ and therefore for higher values of the derivative of S. These simulation results confirm our conjecture stated in the last section that the rate of convergence can be slower

Table 6.2 Hidden information and ARMA learning with recursive prediction error $\rho = 0.9$; $\sigma_u = \sigma_v = 0.1$; $d = 0$

λ	δ t = 500 to 2000	δ t = 2000 to 10,000	Eigenvalues of P_β in real part
0.1	0.389	0.486	–0.241, –0.342
0.145	0.389	0.487	–0.27
0.19	0.388	0.489	–0.269
0.235	0.387	0.490	–0.25
0.28	0.386	0.491	–0.259
0.325	0.385	0.492	–0.252
0.37	0.384	0.493	–0.246
0.415	0.383	0.495	–0.241
0.46	0.381	0.497	–0.235
0.505	0.380	0.499	–0.237
0.55	0.379	0.500	–0.242
0.64	0.377	0.504	–0.261
0.685	0.376	0.506	–0.280
0.73	0.375	0.508	–0.305
0.775	0.373	0.510	–0.343
0.82	0.371	0.513	–0.399
0.865	0.368	0.515	–0.486
0.91	0.365	0.516	–0.405, –0.818
0.955	0.368	0.481	–0.287, –1.37

than 1/2 in least squares learning models when the Benveniste et al. theorem does not apply, and that the higher the derivative of S the lower the rate of convergence. It also confirms that the upper bounds in Mohr (1990) can be reached for ρ close to 1.

Notice that, in table 6.3, the Benveniste et al. theorem applies for $\lambda < 0.595$, but the rates of convergence are much smaller than 1/2 even for sample sizes of 10,000. This shows that the larger the derivative of S the longer it takes for the asymptotic distribution to take over; in other words, for $\lambda = 0.1$ the rate is nearly 1/2, but for larger λ we need a much longer sample size.

Table 6.3 Full information: $\rho = 0.9$; $\sigma_u = \sigma_v = 0.1$; $d = 0$

λ	δ t = 500 to 2000	δ t = 2000 to 10,000	Eigenvalues of S_β in real part
0.1	0.363	0.493	0.09
0.145	0.354	0.488	0.13
0.19	0.345	0.482	0.17
0.235	0.334	0.476	0.21
0.28	0.323	0.468	0.25
0.325	0.310	0.459	0.29
0.37	0.297	0.449	0.33
0.415	0.282	0.435	0.37
0.46	0.268	0.421	0.41
0.505	0.251	0.404	0.45
0.55	0.234	0.386	0.49
0.595	0.216	0.367	0.54
0.64	0.197	0.343	0.58
0.685	0.176	0.319	0.62
0.73	0.156	0.292	0.66
0.775	0.134	0.264	0.70
0.82	0.112	0.234	0.74
0.865	0.089	0.202	0.78
0.91	0.065	0.169	0.82
0.955	0.040	0.136	0.86

The intuition for the slower speed of convergence when the derivative of S is close to unity is straightforward. The least squares learning algorithm adjusts each parameter towards the truth when new information is received (see Marcet and Sargent, 1989a); more precisely, the new belief β_{t+1} will be an average of the previous beliefs β_t and the truth $S(\beta_t)$ plus an error; now, as figure 6.8 shows, if the derivative of S is low $S(\beta_t)$ is very close to β_f itself, while if the derivative is close to unity (as in figure 6.9) $S(\beta_t)$ is close to β_t instead of being close to β_f, so the average can stay far from the fixed point for a long time.

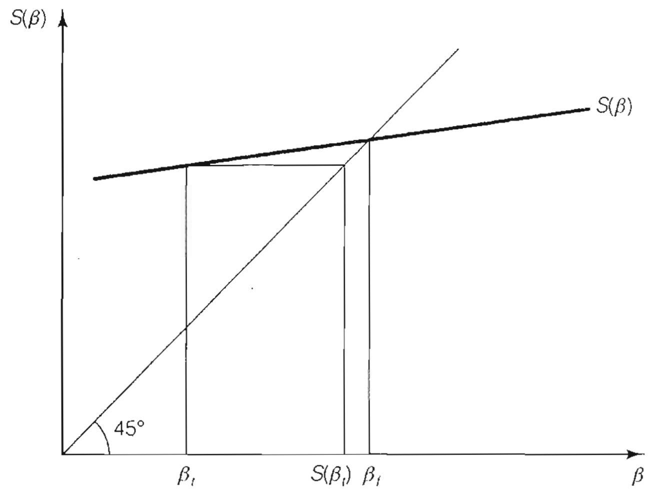

Figure 17 A flat $S(\beta)$ mapping

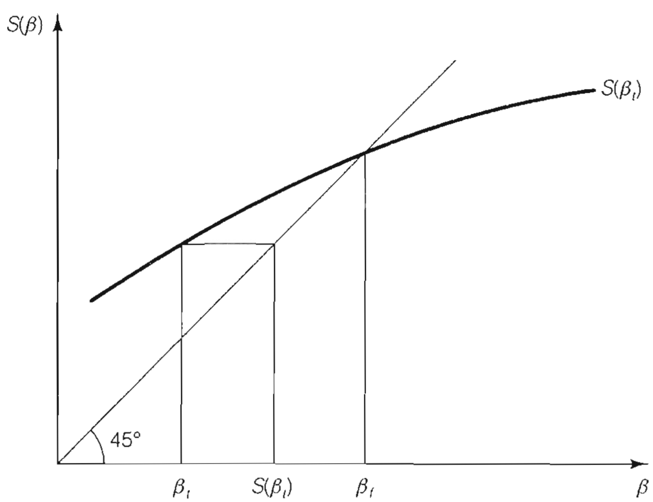

Figure 6.9 A steep $S(\beta)$ mapping

Comparing the results in table 6.1 with those in table 6.3 is of independent interest because they show that, in this model, the rate of convergence is slower with full information than with private information. More precisely, for high values of λ and ρ we have root-t convergence with private information but we have very slow convergence with full information. In this sense, the model with hidden information is more stable than with full information. In the model with full information, even for very large samples, the beliefs have not converged. This means that agents pay a lot of attention to new information that is being received, and that the economy may be moving towards the limit for still quite a while in the full information case, while with hidden information the economy takes fewer periods to converge to its limit.

5 Conclusions

This chapter has described two main extensions to our earlier work on convergence of least squares learning schemes to rational expectations equilibria. First, we showed by example how economic models in which agents are estimating ARMA models can be analyzed using the ordinary differential equations approach. Second, we have obtained some results on the rate of convergence.

Using analytic results from Benveniste et al. and some numerical results from Monte Carlo simulations, we have argued that the speed of convergence to the limiting rational expectations equilibrium is slowed down by higher eigenvalues of the derivative of S at the fixed point. This affects even the rate of convergence; in particular, if the eigenvalues of the derivative of S at the fixed point are larger than 1/2, the speed of convergence is lower than $t^{0.5}$, so that we do not obtain the usual asymptotic distribution in classical econometrics with stationary stochastic processes. Convergence to rational expectations can thus be quite slow, depending mainly on the derivative of the mapping from perceived to actual expectations S.

In the model of this chapter, this leads to the surprising conclusion that it takes a longer time to converge to the rational expectations equilibrium when agents have full information than when agents have hidden information. This happens because the mapping S from believed to actual expectations is much more informative about the fixed point with hidden information.

Also, this low speed of convergence opens up the possibility of having the wrong asymptotic distribution for test statistics when the null hypothesis is rational expectations but the observations are generated by least squares learning. More precisely, any parameter estimate that is a function of β_t may converge to its limiting value at a rate slower than $t^{0.5}$, so that the confidence intervals from classical econometrics will not be correct; in fact, their size will be arbitrarily smaller than the size of the correct intervals as the number of observations goes to infinity. Then assuming rational expectations will lead us to reject the null hypothesis too often, even if the structure of the model economy (leaving aside expectations) is correct. A similar point is made by Bossaerts (1992).

Appendix

In this appendix, we state convergence propositions for the recursive prediction error method and for pseudo-linear regression.

We define the following sets:

D_s $= \{\beta \mid$ the operators $T(\beta)$ and $V(\beta)$ are well defined, and the eigenvalues of $T(\beta)$ are less than unity in modulus $\}$

D_{AS} is the domain of attraction of a fixed point β_f of the ordinary differential equation (6.28)

D_{AP} is the domain of attraction of a fixed point β_f of the ordinary differential equation (6.30)

For the purpose of defining "projection facility" in terms of which a convergence theorem can be stated, we introduce the following additional notation.

$$\begin{aligned} \bar{R}_t &= R_{t-1} + \gamma_t\left(\psi_t \psi_t' - R_{t-1}\right) \\ \bar{\beta}_t' &= \beta_{t-1} + \gamma_t R_t^{-1} \psi_t \hat{w}_t \end{aligned} \tag{A.1}$$

where we recall that $\beta_t = [a_t \;\; c_t]$. Define two sets D_1 and D_2 that satisfy $D_2 \subset D_1 \subset \mathbb{R}^5$. The set D_1 will play the role of a set within which we force the algorithms to stay. In particular, we consider the following modified version of our algorithms:

$$(\beta_t, R_t) = \begin{cases} \bar{\beta}_t, \bar{R}_t & \text{if } (\bar{\beta}_t, \bar{R}_t) \in D_1 \\ \text{some value in } D_2 & \text{otherwise} \end{cases} \tag{A.2}$$

The two propositions stated below pertain to versions of the algorithms (6.24) described in the text that have been modified according to (A.2).

We are free to choose D_2 to be a set that is contained within but is arbitrarily close to D_1. As a practical matter, then, the modified algorithm is defined by the choice of the set D_1.

We make the following assumptions.

Assumption 1 The operator S has a unique fixed point $\beta_f = S(\beta_f)$ that satisfies $\beta \in D_s$.

Assumption 2 For $\beta \in D_s$, T is twice differentiable and V has one derivative.

Assumption 3a The covariance matrix $M_{z_a}(\beta_f)$ is nonsingular.

or

Assumption 3b The covariance matrix $M_{\psi}(\beta_f)$ is nonsingular.

Assumption 4 The process ε_t is serially independent; $E|\varepsilon_t|^p < \infty$ for all $p > 1$.

Assumption 5 There exists a subset Ω_0 of the sample space with $P(\Omega_0) = 1$, two random variables $C_1(\omega)$ and $C_2(\omega)$, and a subsequence $\{t_h(\omega)\}$ for which

$$\left|Z_{t_h}(w)\right| < C_1(w)$$
$$\left|R_{t_h}(w)\right| < C_2(w)$$

for all $\omega \in \Omega_0$ and $h = 1, 2, \ldots$.

Assumption 6 Assume that D_2 is closed, that D_1 is open and bounded, and that $\beta \in D_s$ for all $(\beta, R) \in D_1$. Assume that the trajectories of the ordinary differential equation (6.28) or (6.30) with initial conditions $(\beta(0), R(0)) \in D_2$ never leave a closed subset of D_1.

We now state proposition A1.

Proposition A1 Assume that (β_t, R_t, z_t) are determined via (6.24) as modified by (6.24b′) and (A.2). Suppose that assumptions 1, 2, 3b, and 4 are satisfied.

(i) Assume also that assumptions 5 and 6 are satisfied and that $D_1 \subset D_{AP}$. Then $P(\beta_t \rightarrow \beta_f) = 1$.

(ii) Let $\hat{\beta} \neq \beta_t$, and assume that $M_{\psi}(\beta_f)$ is positive definite. Then $P(\beta_t \rightarrow \hat{\beta}) = 0$.

(iii) If $\mathcal{M}_P$ has one or more eigenvalues with strictly positive real part then $P(\beta_t \rightarrow \beta) = 0$.

For pseudo-linear regression, we have proposition A2.

Proposition A2 Assume that (β_t, R_t, z_t) are determined via (6.24) as modified by (A.2). Suppose that assumptions 1, 2, 3a, and 4 are satisfied.

(i) Assume also that assumptions 5 and 6 are satisfied and that $D_1 \subset D_{AS}$. Then $P(\beta_t \rightarrow \beta_f) = 1$.

(ii) Let $\hat{\beta} \neq \beta_f$ and assume that $M_{z_a}(\beta_f)$ is positive definite. Then $P(\beta_t \rightarrow \hat{\beta}) = 0$.

(iii) If $\mathcal{M}_S$ has one or more eigenvalues with positive real part, then $P(\beta_t \rightarrow \beta_f) = 0$.

These two propositions can be proved simply by retracing the steps of propositions 1, 2, and 3 of Marcet and Sargent (1989a).

Notes

The research on the general subject of this paper has been supported by grants from the National Science Foundation to Carnegie-Mellon University and to the National Bureau of Economic Research. We would like to thank Seppo Honkapohja for telling us about the book by Benveniste, Métivier, and Priouret; Anna Espinal for her research assistance; and Peter Bossaerts for his comments.

1 See also the model of Singleton (1987).

2 For hyperinflation models, this seems a useful assumption. At least during some of the hyperinflations, it is difficult to believe that agents had access to an information set including the history of money supplies.

3 When applied in a "standard" (by which we mean non-self-referential) setting, the recursive prediction error method is known to be statistically consistent and asymptotically efficient. Pseudo-linear regression may or may not be consistent, depending on the parameter values of the ARMA process, but is generally not asymptotically efficient. See Ljung and Söderström (1983, chs 3 and 4) for descriptions of the conditions under which pseudo-linear regressions fail to converge as sample size grows without bound.

4 Also see Kuan (1989). Kuan and White (1991) is a useful treatment of issues related to those studied in this chapter.

5 The arguments of this chapter will extend to higher order systems (i.e. systems with more state variables).

6 The models of Townsend (1983) and Lucas (1975) are examples.

7 At its rational expectations equilibrium, the model of Fourgeaud, Gouriéroux, and Pradel is a version of Sargent and Wallace's (1973) adaptation of Cagan's (1956) model.

8 For example, by a "doubling algorithm" described by Hansen and Sargent (1990).

9 This section is focused on some technicalities which can probably be skipped on a first reading of the chapter. In the computations described in subsequent sections, we always assume that the existence conditions described in this section are satisfied.

10 See, for example, Marcet and Sargent (1989b).

11 See Stoica et al. (1985) for a discussion of a recursive optimal instrumental variable estimator. See Hansen and Sargent (1982) for a treatment of non recursive optimal instrumental variables estimators for a class of linear rational expectations models.

12 To solve the large ordinary differential equation for the recursive prediction error method requires a formula for $E\psi_t\psi_t'$ evaluated at a fixed β. Here is such a formula. Form the stacked state space system

$$\begin{bmatrix} z_{t+1} \\ \psi_{t+1} \end{bmatrix} = \begin{bmatrix} T(\beta) & 0 \\ e_a & -cI \end{bmatrix} \begin{bmatrix} z_t \\ \psi_t \end{bmatrix} + \begin{bmatrix} V(\beta) \\ 0 \end{bmatrix} \varepsilon_{t+1} \qquad (*)$$

or

$$X_{t+1} = H(\beta)X_t + G(\beta)\varepsilon_{t+1} \qquad (\dagger)$$

where

$$X_t = \begin{bmatrix} z_t \\ \psi_t \end{bmatrix}$$

In $(*)$, $T(\beta)$ is 4×4, e_a is 2×4, and $-cI$ is 2×2. The discrete Lyapunov equation for $(\dagger)$ is

$$M_x(\beta) = H(\beta)M_x(\beta)H(\beta)' + G(\beta)\Omega G(\beta)' \qquad (\ddagger)$$

where $\Omega = E\varepsilon_t\varepsilon_t'$. We solve $(\ddagger)$ and pick off the 2×2 matrix on the lower right of this equation to get $E\psi_t\psi_t'$.

13 We used the MATLAB program ode45.m.

14 We did not employ a projection facility in this simulation.

15 Ljung and Söderström (1983) point out that asymptotic distribution results for off-line estimators are only available if they mimic Gauss–Newton algorithms. In the case of maximum likelihood, the asymptotic distribution for the off-line algorithm coincides with the usual distribution of maximum likelihood estimators. For pseudo-linear regressions of exogenous ARMA models, where the direction of the estimator is not updated in the steepest direction to maximize the likelihood function, even though they are consistent, "No explicit expression for the asymptotic covariance matrix for the estimates . . . is known in general" (page 142).

16 Notice that the results for the Gauss–Newton algorithm in Ljung and Söderström (1983) that we mention in the previous note are a special case of this theorem, since the derivative of h in Gauss–Newton algorithms is zero at the true parameters.

17 Some recent results in the learning literature in economics point to similar conclusions. Vives (1993) obtains slower than root-t convergence rates in a model with Bayesian learning, and Mohr (1990) gives a lower bound for the speed of convergence in a simple model, and this bound can be lower than 1/2. Theorem 2.2 in Mohr (1990) is not explicitly presented as a lower bound, but application of the Benveniste et al. theorem that we have described confirms that Mohr only provides upper bounds, since his lower bound is given by λ in our full information model but the derivative of $h(\cdot)$ is $\lambda\rho$. Our simulation results in the next section indicate that those upper bounds are tight when λ is close to 1.

18 In tables 6.1 and 6.2, sometimes one number is recorded in the column labeled eigenvalues, and sometimes two numbers are recorded. When only one number is recorded, it means that the relevant eigenvalues occur as a complex conjugate pair and that we are reporting the pair's common real part.

19 We analyzed learning within a version of this model in Marcet and Sargent (1992).

References

Benveniste, A., Métivier, M. and Priouret, P. 1990: *Adaptive Algorithms and Stochastic Approximations*. Berlin: Springer.

Bossaerts, P. 1994: Asset prices in a speculative market. *Econometric Theory*, forthcoming.

Bray, M. 1982: Learning, estimation, and stability of rational expectations. *Journal of Economic Theory*, 26, 318–39.

Bray, M. 1983: Convergence to rational expectations equilibrium. In Roman Frydman and Edmund Phelps (eds) *Individual Forecasts and Aggregate Outcomes*, Cambridge: Cambridge University Press.

Bray, M. and Savin, N.E. 1986: Rational expectations equilibria, learning, and model specification. *Econometrica*, 54, 1129–60.

Cagan, P. 1956: The monetary dynamics of hyperinflation. In Milton Friedman (ed.), *Studies in the Quantity Theory of Money*, Chicago, IL: University of Chicago Press.

Evans, G. and Honkapohja, S. 1990: Learning, convergence, and stability with multiple rational expectations equilibria. Working Paper, London School of Economics and Political Science.

Fourgeaud, C., Gouriéroux, C. and Pradel, J. 1986: Learning procedure and convergence to rationality. *Econometrica*, 54, 845–68.

Futia, C.A. 1981: Rational expectations in stationary linear models. *Econometrica*, 49, 171–92.

Hansen, L.P. and Sargent, T.J. 1982: Instrumental variables procedures for estimating linear rational expectations models. *Journal of Monetary Economics*, 9 (3), 263–96.

Hansen, L.P. and Sargent, T.J. 1990: Recursive model of dynamic linear models. Manuscript, to be published by Princeton University Press.

Kuan, C.-M. 1989: Estimation of neural network models. Ph.D. thesis, Department of Economics, University of California at San Diego.

Kuan, C.-M. and White, H. 1991: Strong convergence of recursive M-estimators for models with dynamic latent variables. Manuscript, University of Illinois, March.

Ljung, L. and Söderström, T. 1983: *Theory and Practice of Recursive Identification*. Boston, MA: MIT Press.

Ljung, L. 1977: Analysis of recursive stochastic algorithms. *IEEE Transactions on Automatic Control*, AC-22, 551–7.

Lucas, R.E., Jr, 1975: An equilibrium model of the business cycle. *Journal of Political Economy*, 83, 1113–44.

Marcet, A. and Sargent, T.J. 1989a: Convergee of least squares learning mechanisms in self referential linear stochastic models. *Journal of Economic Theory*, 48 (2), 337–68.

Marcet, A. and Sargent, T.J. 1989b: Convergence of least squares learning in environments with hidden state variables and private information. *Journal of Political Economy*, 97(6), 1306–22.

Marcet, A. and Sargent, T.J. 1992: The convergence of vector autoregresions to rational expectations equilibrium. In Alessandro Vercelli and Nicola Dimitri (eds), *Macroeconomics: A Strategic Survey*, Oxford: Oxford University Press, 139–64.

Mohr, M. 1990: Asymptotic theory for least squares estimators in regression models with forecast feedback. Working Paper, Bonn University.

Muth, J.F. 1960: Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association*, 55 (290), 299–306.

Sargent, T.J. 1991: Equilibrium with signal extraction from endogenous variables. *Journal of Economic Dynamics and Control*, 15 (2), 245–74.

Sargent, T.J. and Wallace, N. 1973: Rational expectations and the dynamics of hyperinflation. *International Economic Review*, 63 (3), 328–50.

Singleton, K.J. 1987: Asset prices in a time series model with disparately informed, competitive traders. In William A. Barnett and Kenneth J. Singleton (eds), *New Approaches to Monetary Economics*, Cambridge: Cambridge University Press.

Stoica, P., Söderström, T. and Friedlander, B. 1985: Optimal instrumental variable estimates of AR parameters of an ARMA process. *IEEE Transactions on Automatic Control*, AC-30, 1066–74.

Townsend, R.M. 1983: Forecasting the forecasts of others. *Journal of Political Economy*, 91 (4), 546–88.

Vives, X. 1993: How fast do rational agents learn? *Review of Economic Studies*, April, 329–47.