Properties of Rational Expectations Equilibrium with an Adaptive Learning Approach [*†]

Yuanying Guan

February 13, 2009

Abstract

In this paper, I will present the central idea along with some results from Evans and Honkapohja [6] concerning the convergence of a recursive learning scheme to rational expectations equilibrium solutions in macroeconomic models. The unified modeling framework is centered around a stability condition known as 'expectational stability'. To obtain this condition, stochastic approximation is applied to associate a recursive learning algorithm with an ordinary differential equation. I will show a basic convergence theorem in detail in this stochastic approximation approach (Ljung [9]). Then, I will apply this theorem to the Cobweb model under certain assumptions.

Keywords rational expectations equilibrium; expectational stability; stochastic recursive learning; locally asymptotically stable; stochastic approximation

[*] This paper is being submitted to the Department of Mathematics for the Advanced Topics Exam in Financial Mathematics.

[†] Academic Advisors: Dr. Alec Kercheval and Dr. Paul Beaumont

Contents

1 Introduction

Contemporary macroeconomics gives due weight to the role of expectations. A central aspect is that expectations influence the time path of the economy, and one might reasonably hypothesize that the time path of the economy influences expectations. The current standard methodology for modeling expectations is to assume *rational expectations* (RE), which is in fact an equilibrium in this two-sided relationship. Formally, in dynamic stochastic models, rational expectations is usually defined as the mathematical conditional expectation of the relevant variables. The expectations are conditioned on all of the information available to the decision makers. Rational expectations theory defines this kind of expectations as being identical to the *best guess of the future* that uses all available information. As a result, rational expectations do not differ systematically or predictably from equilibrium results.

We emphasize that rational expectations is in fact an equilibrium concept. The actual stochastic process followed by prices depends on the forecast rules used by agents, so that the optimal choice of the forecast rule by any agent is conditional on the choices of others. An RE equilibrium imposes the consistency condition that each agent's choice is a best response to the choices by others. In the simplest models we have representative agent. Economists use the term representative agent to refer to the typical individual of a certain type (for example, the typical consumer, or the typical firm). More technically, an economic model is said to have a representative agent if all agents of the same type are identical. Thus, in our simplest models, all those choices made by agents are identical.

Based on need for our future research, we investigate properties of *rational expectations equilibrium* (REE) in this paper. Specifically, we present some results concerning the topic of convergence of adaptive learning to REE solutions by Evans and Honkapohja [6]. They have done an excellent job motivating, applying, and generalizing the E-stability condition, which gives us a very good criteria to recognize 'learnable' REEs.

The rest of this paper is organized as follows. Section 2 will describe the Lucas asset pricing model in some detail. Section 3 presents some results from Evans and Honkapohja [6] concerning the convergence of recursive learning scheme to REE solutions. In particular, a widely recited basic theorem in stochastic approximation is shown (Ljung [9]). Also, an application of this stochastic approximation result to a specific model, the Cobweb model, will be discussed.

2 The Lucas Asset Pricing Model

We begin with the description of Lucas model (Lucas,1978). Suppose there are N heterogeneous, infinitely lived agents and a single risky asset (stock) paying stochastic dividends according to a distribution the agents are assumed to know. In each period each agent will have a certain wealth, which is the sum of investment gains from the last period and the endowment for current period. He may either consume them during that period or he may use them to invest in next period.

The agents in this model have preferences. These preferences will be presented by a utility function. A utility function will take, as input, the agent's consumption and return a numerical value to indicate the level of happiness achieved from that amount of consumption. The agent's objective will be to maximize their total satisfaction. The agents are endowed at time 0 with initial stock holdings s_i. At time 0, agent i solves the following problem:

$$\max_{\{c_t^i, s_{t+1}^i\}_0^\infty} E\{\sum_{t=0}^{\infty} \beta_i^t u^i(c_t^i)\} \tag{2.1}$$

subject to:

$$c_t^i + p_t s_{t+1}^i \leq s_t^i(p_t + d_t) + e_t^i, \forall t \tag{2.2}$$

$$c_t^i \geq 0, \forall t, s_0^i \ given$$

where β_i is agent i's discount factor, c_t^i and s_t^i are agent i's period t consumption and share holdings, respectively, d_t is the stochastic dividend for period t, p_t is the price of the stock and e_t^i is the period t endowment of agent i.

In asset pricing models, agents are interested to know the market clearing price function based on several parameters. Let us denote the set of all parameters at time t by θ_t. Included in θ_t is the dividend and price, the wealth of the agents (s_t), the distribution of the dividends, and the total supply of the stock (N). Let us denote by θ_t^i the individual state for agent i. We now wish to define a very important equilibrium concept (see Culham [1]).

Definition 2.1: A *Rational Expectations Equilibrium* (REE) is a set of demand functions $\{c^i(\theta_t^i), s^i(\theta_t^i)\}_{i=1}^N$, and a pricing function $p(\theta_t)$ such that the following conditions hold:

1. These functions solve each agent's optimization problem (i.e. the first order conditions hold)
2. All markets clear:

(i) $\Sigma_{i=1}^N s_t^i = N, \ \forall t$

(ii) $\Sigma_{i=1}^N (c_t^i + p_t s_{t+1}^i) = \Sigma_{i=1}^N (s_t^i(p_t + d_t) + e^i), \forall t, \ where \ p_t = p(\theta_t)$

3 Convergence of Recursive Learning to Rational Expectations Equilibrium

In the 1950s and 1960s expectations were introduced into almost every area of macroeconomics, including consumption, investment, money demand, and inflation. Typically, expectations were mechanically incorporated in macroeconomic modeling using adaptive expectations or related lag schemes. *Adaptive expectations* was formally introduced in the 1950s by Cagan (1956), Friedman(1957) and Nerlove (1958). In terms of the price level adaptive expectation takes the form

$$p_t^e = p_{t-1}^e + \lambda(p_{t-1} - p_{t-1}^e)$$

where p_t^e, p_{t-1}^e denote the expectation of price at time t and time $t-1$. λ is between 0 and 1. It can also be written in the form

$$p_t^e = \lambda \sum_{i=0}^{\infty} (1-\lambda)^i p_{t-1-i}.$$

Rational expectations theories were developed in response to perceived flaws in theories based on adaptive expectations. Under adaptive expectations, expectations of the future value of an economic variable are based on past values. For example, people would be assumed to predict inflation by looking at inflation last year and in previous years. If the economy suffers from constantly rising inflation rates (perhaps due to government policies), people would be assumed to always underestimate inflation. This may be regarded as unrealistic - surely rational individuals would sooner or later realize the trend and take it into account in forming their expectations.

The hypothesis of rational expectations addresses this criticism by assuming that individuals take all available information into account in forming expectations. Though expectations may turn out incorrect, they will not deviate systematically from the expected values. Muth (1961) was the first to formulate explicitly the notion of rational expectations and did so in the context of the Cobweb model which we will explain in section 3.1. Rational expectations then made the decisive appearance in macroeconomics in the papers of Lucas (1972) and Sargent (1973).

Under rational expectations assumption, we write:

$$p_t^e = E_{t-1} p_t$$

Here $E_{t-1} p_t$ denotes the mathematical (statistical) expectation of p_t conditional on variables observable at time $t-1$ (including past data).

With the widespread use of rational expectations as the equilibrium concept in dynamic economic models, the issue of convergence of adaptive learning mechanisms to rational expectations equilibria has recently received considerable attention.

Adaptive learning is a new viewpoint of expectations. It emphasizes the 'adaptive' process in agent's learning scheme rather than provides agent adaptive expectations directly. Agents are able to adjust their forecast rule as new data become available over time. Taking this approach immediately raises the question of its relationship to rational expectations. It turns out that in many cases learning can provide at least an asymptotic justification for the RE hypothesis. For example, in the Cobweb model with unobserved *iid* shocks, if agents estimate a constant expected value by computing the sample mean from past prices, one can show that expectations will converge over time to the RE value. This property turns out to be quite general for the Cobweb-type models, provided agents use the appropriate econometric functional form.

Another major advantage of the learning approach arises in connection with the issue of multiple equilibria. In nonlinear models this issue of multiplicity of RE solutions has been frequently encountered. In the adaptive learning approach it is supposed that agents start with initial estimates of the parameters of a stochastic process for p_t and revise their estimates, following standard econometric procedures, as new data points are generated. This enable us to use adaptive learning as a selection criterion, since only one of the RE solutions can emerge in the long run under certain initial conditions.

Most content of section 3 comes from Evans and Honkapohja [6] (see Chapter 1,2,5,6,7 and 11).

3.1 The Cobweb Model

The Cobweb model we introduce here follows a version from Bray and Savin [2]. We consider firms producing a homogeneous good. The firms make their production decisions at each date t before the realization of an exogenous stochastic demand

$$d_t = m_I - m_p p_t + v_{1t}$$

where m_I, m_p are constant, $m_p > 0$, p_t is the market clearing price of good, and v_{1t} is an unobserved white noise demand shock.

Assume that the average supply to market depends positively on the expected price

$$s_t = r_I + r_p p_t^e + r'_\omega \omega_{t-1} + v_{2t}$$

where r_I, r_p are constant, $r_p > 0$, v_{2t} is unobserved white noise shock, ω_{t-1} is a vector of observable shocks, and r_ω is a constant vector.

The market clearing condition $s_t = d_t$ implies that

$$p_t = \mu + \alpha p_t^e + \delta' \omega_{t-1} + \eta_t \tag{3.1}$$

where $\mu = (m_I - r_I)/m_p$, $\delta = -m_p^{-1} r_\omega$, and $\alpha = -r_p/m_p$. Note that $\alpha < 0$, $\eta_t = (v_{1t} - v_{2t})/m_p$, so that we can write $\eta_t \sim$ iid $N(0, \sigma_\eta^2)$. Under rational expectations, $p_t^e = E_{t-1}p_t$, operating with E_{t-1} on both sides of equation (3.1) and solving for $E_{t-1}p_t$, we obtain

$$E_{t-1}p_t = (1-\alpha)^{-1}\mu + (1-\alpha)^{-1}\delta'\omega_{t-1}$$

Since also $p_t - E_{t-1}p_t = \eta_t$, it follows that there is a unique rational expectations equilibrium given by

$$p_t = \overline{a} + \overline{b}'\omega_{t-1} + \eta_t$$

where

$$\overline{a} = (1-\alpha)^{-1}\mu \text{ and } \overline{b} = (1-\alpha)^{-1}\delta$$

Although the REE is unique, we can still ask whether it is *learnable* in the following sense. Suppose that firms believe that prices follow the process

$$p_t = a + b'\omega_{t-1} + \eta_t \tag{3.2}$$

corresponding to the REE, but that a and b are unknown to them. We assume that equation (3.2) is the *perceived law of motion* of the firms (what firms believe about the evolution of the price) and that they attempt to estimate a and b. We suppose that firms estimate a and b by a least square regression of p_t on ω_{t-1} and an intercept. Their estimates will be updated over time as more information is collected. Their forecast at $t-1$ are given by

$$p_t^e = a_{t-1} + b_{t-1}'\omega_{t-1} \tag{3.3}$$

The question of interest is whether $a_t \to \overline{a}$ and $b_t \to \overline{b}$ as $t \to \infty$. We will give out the answer in the following discussion.

3.2 E-stability

The basic required concept for this part is a map from *perceived law of motion* (PLM) to the *actual law of motion* (ALM). Agents will have some perceptions about the evolution of the economy, or more specific, the price in certain models before the realization of this evolution. This is defined as *perceived law of motion* in our following discussion. Correspondingly, the actual evolution of economy (price) is called *actual law of motion.* The *E-stability principle* stated in its most comprehensive form is that the mapping from the PLM to the ALM governs the stability of equilibria under learning. More specifically, E-stability conditions obtained from this mapping provide conditions for asymptotic stability of an REE under least square learning. We focus here on obtaining this condition for the Cobweb model.

We begin with the assumption that agents have a PLM which they use to make forecasts of the variables of interest. Usually we take the form of the PLM that correspond to the REE of interest. Thus in the current case we take the PLM to be of the form $p_t = a + b' w_{t-1} + \eta_t$. For $a = \overline{a}$ and $b = \overline{b}$, the PLM would be the REE, but we allow for the possibility that agents have nonrational expectations. p_t^e is given by

$$p_t^e = a + b'\omega_{t-1} \tag{3.4}$$

Inserting this equation into (3.1), one can solve for the *actual law of motion*, implied by PLM:

$$p_t = (\mu + \alpha a) + (\delta + \alpha b)'\omega_{t-1} + \eta_t \tag{3.5}$$

This implicitly defines the mapping from the PLM to the ALM

$$T\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \mu + \alpha a \\ \delta + \alpha b \end{pmatrix} \tag{3.6}$$

We can now define E-stability in the form appropriate for determining the stability of the REE under least squares learning. Note first that the unique REE for our model is the unique fixed point of the T-map (3.6). Consider the differential equation

$$\frac{d}{d\tau}\begin{pmatrix} a \\ b \end{pmatrix} = T\begin{pmatrix} a \\ b \end{pmatrix} - \begin{pmatrix} a \\ b \end{pmatrix} \tag{3.7}$$

where τ denotes 'notional' or 'artificial' time. We say that the REE is *expectationally stable*, or *E-stable*, if the REE is locally asymptotically stable under equation (3.7). Intuitively, E-stability determines stability of the REE under a stylized rule in which PLM parameters a and b are adjusted slowly in the direction of the implied ALM parameters.

Here we would like to introduce a basic concept concerning differential equations.

Local stability of the autonomous differential equation

$$\frac{d\theta}{d\tau} = f(\theta(\tau)) \tag{3.8}$$

can be determined by the linearization of $f(\theta)$ at an equilibrium point. Here $f : W \to \mathbb{R}^m$, where $W \subset \mathbb{R}^m$, W open.

Definition 3.1. *$\overline{\theta} \in W$ is an equilibrium of equation (3.8) if $f(\overline{\theta}) = 0$.*

Definition 3.2. *$\overline{\theta}$ is said to be locally stable if for every $\varepsilon > 0$, there exists $\delta > 0$ such that $|\theta(\tau) - \overline{\theta}| < \varepsilon$ for all $|\theta(0) - \overline{\theta}| < \delta$. $\overline{\theta}$ is said to be locally asymptotically stable if $\overline{\theta}$ is locally stable and in addition $\lim_{\tau\to\infty} \theta(\tau) \to \overline{\theta}$ for all $\theta(0)$ in some neighborhood of $\overline{\theta}$.*

We have the following result (see Hirsch and Smale [7], Sections 9.1-9.2, p.181, and p.187).

Proposition 3.3. *Assume f is continuously differentiable on W. Let $A = Df(\overline{\theta})$ denote the (Jacobian) derivative of f at $\overline{\theta}$, where $\overline{\theta}$ is an equilibrium of (3.8). If all eigenvalues of A have negative real parts, then $\overline{\theta}$ is locally asymptotically stable with respect to system (3.8). If A has one or more eigenvalues with positive real parts, then $\overline{\theta}$ is unstable with respect to system (3.8).*

To determine E-stability in our example, combine equations (3.6) and (3.7) and write the differential equation component to obtain

$$\frac{da}{d\tau} = \mu + (\alpha - 1)a$$

$$\frac{db_i}{d\tau} = \delta_i + (\alpha - 1)b_i, \ for \ i = 1, \ldots, n$$

where n is the dimension of ω. It follows that the REE is E-stable if and only if $\alpha < 1$.

3.3 Recursive Least Squares

We begin by noting that the standard least squares regression formula has a recursive formulation. In fitting the equation $y_i = c'x_i + e_i$ using data $i = 1, \ldots, T$ on the $k \times 1$ independent vector x_i and the dependent variable y_i, the value of the $k \times 1$ coefficient vector c which minimizes $\sum_{i=1}^{T} e_i^2$ is given by the least squares formula

$$c = (\sum_{i=1}^{T} x_i x_i')^{-1} (\sum_{i=1}^{T} x_i y_i')$$

c can instead be computed using the *recursive least squares* (RLS) formulas

$$c_t = c_{t-1} + t^{-1}R_t^{-1}x_t(y_t - x_t'c_{t-1})$$

$$R_t = R_{t-1} + t^{-1}(x_tx_t' - R_{t-1})$$

c_t and R_t denote the coefficient vectors and the moment matrix for x_t using data $i = 1, \ldots, t$. To generate the least square values, the initial value for the recursion must be set appropriately.[1] With these initial values, equations above generates the usual least squares formula for c_t, the least squares coefficient vector using data $i = 1, \ldots, t$, and c above is given by $c = c_T$.

We now apply the RLS formulas to our learning problem. Our agents are running a least squares regression of p_t on z_{t-1}, where $z_{t-1}' = (1\ w_{t-1}')$. For convenience, write

$$\phi_t = \begin{pmatrix} a_t \\ b_t \end{pmatrix}$$

Applying the RLS formulas, we obtain

$$\phi_t = \phi_{t-1} + t^{-1}R_t^{-1}z_{t-1}(p_t - z_{t-1}'\phi_{t-1})$$

$$R_t = R_{t-1} + t^{-1}(z_{t-1}z_{t-1}' - R_{t-1})$$

Since p_t is given by equations (3.1) and (3.3), we have

$$p_t = (\mu + \alpha a_{t-1}) + (\delta + \alpha b_{t-1})'w_{t-1} + \eta_t \tag{3.9}$$

or

$$p_t = z_{t-1}'T(\phi_{t-1}) + \eta_t \tag{3.10}$$

where $T(\phi) \equiv T\begin{pmatrix} a_t \\ b_t \end{pmatrix}$ is given by equation(3.6). Combining equations, we arrive at the stochastic recursive system

$$\phi_t = \phi_{t-1} + t^{-1}R_t^{-1}z_{t-1}(z_{t-1}'(T(\phi_{t-1}) - \phi_{t-1}) + \eta_t) \tag{3.11}$$

$$R_t = R_{t-1} + t^{-1}(z_{t-1}z_{t-1}' - R_{t-1}) \tag{3.12}$$

We want to know whether equations (3.11)-(3.12) converge as $t \to \infty$. Let $\overline{\phi'} = (\overline{a}, \overline{b'})$. Our claim is that in the case of $\eta_t = 0$, if $\alpha < 1$, then $\phi_t \to \overline{\phi}$ with probability 1. Since $T(\overline{\phi}) = \overline{\phi}$, it also follows from equation(3.10) that the price process converges to the REE.

[1] Assuming $X_k = (x_1, \ldots, x_k)'$ is of full rank and letting y^k denote $y^k = (y_1, \ldots, y_k)'$, the initial value c_k is given by $c_k = (X_k'X_k)^{-1}X_k'y^k$ and the initial value R_k is given by $R_k = k^{-1}X_k'X_k$

To show convergence formally requires results from the stochastic approximation literature. Here we will first show how to write our problem into a standard stochastic recursive system in section 3.4, then we present stochastic approximation technique in section 3.5.

3.4 Convergence of Stochastic Recursive Algorithms

There is a substantial literature in statistics and engineering which concerns itself precisely with the convergence of stochastic recursive algorithms such as equation (3.11)-(3.12). (This method is also called stochastic approximation.) Marcet and Sargent (1989) show how this technique could be applied in economics to the analysis of adaptive learning. In this section and next section we provide the central technique.

We consider a *stochastic recursive algorithm* (SRA) of the form

$$\theta_t = \theta_{t-1} + \gamma_t Q(t, \theta_{t-1}, X_t) \tag{3.13}$$

where θ_t is a vector of parameter estimates, X_t is a random vector defined below, which is in general a function of the previous estimates $\{\theta_1, \theta_2, \ldots, \theta_{t-1}\}$ and a sequence of random shocks W_t. γ_t is a deterministic sequence of 'gains'. The function Q expresses the way in which the estimate θ_{t-1} is revised in line with the last period's observations.

In most economic models we can take the following structure for X_t

$$X_t = A(\theta_{t-1})X_{t-1} + B(\theta_{t-1})W_t \tag{3.14}$$

where W_t is a random process. In particular, there is a possibility to treat the sequence W_t either in a stochastic or in a deterministic framework.

In our Cobweb example, we could do even more. We are able to write X_t in the following form to simplify the whole process

$$X_t = AX_{t-1} + BW_t \tag{3.15}$$

which means we treat X_t as an independent random vector.

We give the details here of how equations (3.11)-(3.12) can be put into the standard SRA form (3.13). Also, we add another restriction to the Cobweb model in form of (3.15) so that we complete the dynamics of the whole system.

We would like to define θ_t to include all the components of ϕ_t and R_t. To do this we define $S_{t-1} = R_t$. The system can be rewritten as

$$\phi_t = \phi_{t-1} + t^{-1} S_{t-1}^{-1} z_{t-1} (z'_{t-1}(T(\phi_{t-1}) - \phi_{t-1}) + \eta_t)$$

$$S_t = S_{t-1} + t^{-1}(\frac{t}{t+1})(z_t z'_t - S_{t-1})$$

This system is now implicitly in standard form (3.13) with the following definitions of variables:

$$\theta_t = vec(\phi_t \; S_t)$$

$$\gamma_t = t^{-1}$$

$$X_t = \begin{pmatrix} 1 \\ \omega_t \\ \omega_{t-1} \\ \eta_t \end{pmatrix}$$

Recall that $z'_t = (1 \; \omega'_t)$. Thus all the components of z_t and z_{t-1} have been included in X_t. Here *vec* denotes the matrix operator which stacks in order the columns of matrix $(\phi_t \; S_t)$ into a column vector. Now we have

$$\theta_t = \theta_{t-1} + \gamma_t \begin{pmatrix} Q_\phi(t, \theta_{t-1}, X_t) \\ Q_s(t, \theta_{t-1}, X_t) \end{pmatrix} \tag{3.16}$$

The function $Q(t, \theta_{t-1}, X_t)$ is given by

$$Q_\phi(t, \theta_{t-1}, X_t) = S_{t-1}^{-1} z_{t-1} (z'_{t-1}(T(\phi_{t-1}) - \phi_{t-1}) + \eta_t)$$

$$Q_s(t, \theta_{t-1}, X_t) = vec((\frac{t}{t+1})(z_t z'_t - S_{t-1}))$$

Now we define the dynamics of X_t for the Cobweb model. We borrow the assumption on observable shocks w_t from the Multivariate Muth Model (see Evans and Honkapohja [6], Section 6.6, p.141-144).

$$w_t = C w_{t-1} + v_t \tag{3.17}$$

here w_t is an $n \times 1$ vector of observable shocks, v_t is an $n \times 1$ vector of white noise shocks. C is assumed to be an $n \times n$ known matrix, whose eigenvalues lie inside the unit circle.

So we could write X_t as

$$X_t = A X_{t-1} + B W_t$$

where

$$A = \begin{pmatrix} 0\,0\,0\,0 \\ 0\,C\,0\,0 \\ 0\,0\,C\,0 \\ 0\,0\,0\,0 \end{pmatrix}, B = \begin{pmatrix} 1\,0\,0\,0 \\ 0\,I\,0\,0 \\ 0\,0\,I\,0 \\ 0\,0\,0\,1 \end{pmatrix}, W_t = \begin{pmatrix} 1 \\ v_t \\ v_{t-1} \\ \eta_t \end{pmatrix}$$

Further more, assume that C be a zero matrix, we could write

$$\omega_t = v_t$$

then

$$X_t = BW_t = W_t \tag{3.18}$$

A more general form of X_t requires more work in verification of the assumptions of our main theorem, but the idea does not change by this simplicity. Thus, this simple form of X_t is applied to our further discussion for the Cobweb model. Having shown that the system can be placed in standard SRA form, the next step is to present the convergence result we need.

The stochastic approximation results show that the behavior of the SRA is well approximated by the behavior of the associated ODE for large t. In particular, we will present a basic result in a widely cited paper written by Ljung [9] and provide the outline of proof in section 3.5. Also, we will write out a rigorous proof of this general theorem with simplified form (3.18) of X_t in Appendix B.

3.5 Stochastic Approximation

The stochastic approximation approach associates an ordinary differential equation (ODE) with the SRA,

$$\frac{d\theta}{d\tau} = h(\theta(\tau)) \tag{3.19}$$

where $h(\theta)$ is obtained as

$$h(\theta) = \lim_{t\to\infty} EQ(t, \theta, X_t(\theta)) \tag{3.20}$$

provided this limit exists. E denote the expectation of $Q(t, \theta, X_t(\theta))$, for θ fixed.

It's necessary to formulate the assumptions on the learning rules. Suppose we have

$$\theta_t = \theta_{t-1} + \frac{1}{t}Q(t, \theta_{t-1}, X_t) \tag{3.21}$$

$$X_t = AX_{t-1} + BW_t \tag{3.22}$$

where A has all eigenvalues strictly inside the unit circle, B is a bounded matrix. Let

$$D_s = \mathcal{R}^n \tag{3.23}$$

given that the dimension of θ is n. Then there exits a λ, such that

$$|A^k| < C\lambda^k; \quad \lambda < 1 \tag{3.24}$$

according to A has all eigenvalues strictly inside the unit circle. The regularity conditions below will be assumed to be valid in D_s. Now, we postulate the following assumptions:

(A.1) W_t is a sequence of independent random variables and $|W_t| < C$ with probability 1 at all t.

Note that (A.1) and the requirements for matrix A and B imply the existence of a bounded invariant set D_X, such that $X_0 \in D_X$ implies $X_t \in D_X$ for all t, with probability 1. We also have the following assumption

(A.2) The function $Q(t, \theta, X_t)$ has a bounded first order derivative w.r.t θ for $\theta \in D_R$ and $X_t \in D_X$.

(A.3) The limit $\lim_{t\to\infty} EQ(t, \theta, X_t) = h(\theta)$ exists for $\theta \in D_R$

(A.4) Define $q_{t,\theta} = \frac{1}{t}\sum_{s=1}^{t} Q(s, \theta, X_s)$, assume

$$\lim_{t\to\infty} q_{t,\theta} = \lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} Q(s, \theta, X_s) = h(\theta)$$

Here comes our main theorem

Theorem 3.4: *Consider the algorithm (3.21),(3.22) subject to assumptions (A.1)-(A.4). Assume that*
1) $\overline{D}$ is a compact subset of $\mathcal{R}^n$ such that the trajectories of (3.19) and (3.21) that start in $\overline{D}$ remain in a compact subset of $\mathcal{R}^n$ for $\tau > 0$.
2) the differential equation (3.19) has an invariant set D_c with domain of attraction $D_A \supset \overline{D}$. Then $\theta_t \to D_c$ with probability 1 as $t \to \infty$.

Remarks: An invariant set of a differential equation is a set such that the trajectories remain there for $-\infty < \tau < \infty$. The domain of attraction of an invariant set D_c consists of those points from which the trajectories converge into D_c as τ tend to infinity. An interesting special case is when D_c is just a locally stable equilibrium point of (3.19) say $\overline{\theta}$. And we will choose $D_c = \overline{\theta}$ in the proof shown below.

We need the following Lemma to prove Theorem 3.4. (see Appendix A for complete proof)

Lemma 3.5: *Consider the algorithm (3.21), (3.22) under assumptions (A.1)-(A.4). Let $\theta^* \in \overline{D}$ be a cluster point of θ_t for fixed ω, let $\theta_{t_k} \to \theta^*$ be a convergent subsequence of θ_t for fixed ω. Assume that $\theta^* \neq \overline{\theta}$, where $\overline{\theta}$ represents the equilibrium point of the associated ODE of algorithm (3.21),(3.22).*

Then there exists $\rho(\theta^*)$ such that if $\theta_{t_k} \in \mathcal{B}(\theta^*, \rho(\theta^*))$ for $\forall k$, there exists $\Delta\tau_0 = \Delta\tau_0(\theta^*, \rho(\theta^*))$ and $\epsilon_0(\rho)$ such that for $\forall \Delta\tau < \Delta\tau_0$, there exists a sequence $\{m_{t_k,\Delta\tau}\}$, satisfying that $m_{t_k,\Delta\tau} > t_k$, and that

$$|\sum_{t=t_k}^{m_{t_k,\Delta\tau}} \frac{1}{t} - \Delta\tau| < \epsilon_0(\rho)\Delta\tau, \quad \textit{when} \quad k \to \infty$$

There exist $q^1_{t_k,m_k,\theta^*}, q^2_{t_k,m_k,\theta^*}$, such that

$$\theta_{m_k} = \theta_{t_k} + \Delta\tau h(\theta^*) + q^1_{t_k,m_k,\theta^*} + q^2_{t_k,m_k,\theta^*} \tag{3.25}$$

where $m_k = m_{t_k,\Delta\tau}$ and

$$q^1_{t_k,m_k,\theta^*} \to 0 \textit{ as } t \to \infty \tag{3.26}$$

and

$$|q^2_{t_k,m_k,\theta^*}| \leq 4\Delta\tau\mathcal{K}\rho \tag{3.27}$$

where $\mathcal{K}$ denotes the bound of the first order partial derivative of $Q(t,\theta,X)$ with respect to θ.

Outline of proof: The basic idea of the proof is intuitive.
An explicit expression for θ_{j_k} where

$$t_k \leq j_k \leq m_{t_k,\Delta\tau}$$

is derived. This expression shows that if certain terms are small, then θ_{j_k} is close to what would be obtained if $Q(s,\theta_{s-1},X_s), t_k < s \leq j_k$ is replaced by $h(\theta^*)$. Then we show that those terms in above expression are indeed small. A complication so far has been that in order to prove these things it must be assumed that $\theta_{m_{t_k,\Delta\tau}}$ remains in a small neighborhood of θ^*. It is proved by induction that $\theta_s, t_k < s \leq m_{t_k,\Delta\tau}$ actually will remain in this neighborhood up to $m_{t_k,\Delta\tau}$ if $\Delta\tau$ is chosen sufficiently small and k sufficiently large. This will conclude the proof of Lemma 3.5.

It follows from the converse stability theorems (see Krasovskij [8]) that the stability assumption for $D_c = \overline{\theta}$ in Theorem 3.4 implies the existence of a function V (Lyapunov function) with properties

1) $V(\theta)$ is infinitely differentiable
2) $0 \leq V(\theta) < 1$ for $\theta \in D_A$ and $V(\theta) = 0 \Longleftrightarrow \theta = \overline{\theta}$
3) $(d/d\tau)V(\theta(\tau)) = V'(\theta)h(\theta) \leq 0$ for $\theta \in D_A$ and equality holds only for $\theta = \overline{\theta}$

We now proceed with proof of the main theorem. A rigorous proof is given in Appendix B, an outline of the proof is as follows:

Outline of Proof: Consider a convergent subsequence θ_{t_k} of θ_t for any fixed ω, let θ^* be the limit of θ_{t_k} and $\theta^* \neq \overline{\theta}$. Then θ_{t_k} is close enough to θ^* when $k \to \infty$. According to Lemma 3.5, $\theta_{m_{t_k,\Delta\tau}}$ will approximately be $\theta_{t_k} + \Delta\tau h(\theta^*)$. This means that $V(\theta_{m_{t_k,\Delta\tau}})$ is strictly less than $V(\theta_{t_k})$ if we carefully choose certain quantities. By showing that $\lim_{t\to\infty} \inf V(\theta_t) = 0$, θ^* is proved to be a cluster point of θ_t. Then $\lim_{t\to\infty} \sup V(\theta_t) = 0$ implies that θ^* is the only possible cluster point for θ_t. The formal proof is somewhat lengthy and involves several elaborate choices of constants.

3.6 Application to the Cobweb model

In this part we show how to apply the stochastic approximation results in section 3.5 to the special case of the Cobweb model where $\eta_t = 0$ to obtain the stability result. First, we need to make sure the algorithm for the Cobweb model satisfies assumptions (A.1)-(A.4).

For the Cobweb model, as we presented in section 3.4, we have

$$\theta_t = \theta_{t-1} + \gamma_t \begin{pmatrix} Q_\phi(t, \theta_{t-1}, X_t) \\ Q_s(t, \theta_{t-1}, X_t) \end{pmatrix} \tag{3.28}$$

where

$$Q_\phi(t, \theta_{t-1}, X_t) = S_{t-1}^{-1} z_{t-1}(z_{t-1}'(T(\phi_{t-1}) - \phi_{t-1}) + \eta_t)$$
$$Q_s(t, \theta_{t-1}, X_t) = vec((\frac{t}{t+1})(z_t z_t' - S_{t-1}))$$

and

$$X_t = W_t \tag{3.29}$$

where W_t is bounded w.p.1 and X_t is bounded consequently.

(A.1) W_t is a sequence of independent random variables and $|W_t| < C$ with probability 1 at all t.

We have $W_t = \begin{pmatrix} 1 \\ v_t \\ v_{t-1} \\ \eta_t \end{pmatrix}$ in the Cobweb model, we assume v_t, η_t be independent, bounded noises. Now the model satisfy both (A.1) and the requirements for matrix A and B, since that A is zero matrix and B is identity matrix.

(A.2) The function $Q(t, \theta, X_t)$ has a bounded first order partial derivative w.r.t θ for $\theta \in D_R$ and $X_t \in D_X$.

Since

$$Q_\phi(t, \theta_{t-1}, X_t) = S_{t-1}^{-1} z_{t-1}(z'_{t-1}(T(\phi_{t-1}) - \phi_{t-1}) + \eta_t)$$

$$Q_s(t, \theta_{t-1}, X_t) = vec((\frac{t}{t+1})(z_t z'_t - S_{t-1}))$$

Given that S_t^{-1} is bounded, it would be easy to say Q has a bounded first order partial derivative w.r.t θ since all other parts are about $\{\omega_t\}$ (which are bounded) and $T'(\phi_{t-1}) - \phi'_{t-1}$ (which is linear thus a constant when applying derivative). Here, an assumption is needed that S_t^{-1} exists for $\forall t$. But that is reasonable since our initial value $S_0 = R_1$ is a full rank, symmetric matrix. Also, we know that

$$S_t = S_{t-1} + t^{-1}(\frac{t}{t+1})(z_t z'_t - S_{t-1}) = \frac{1}{t+1} S_0 + \frac{1}{t+1}(z_t z'_t + z_{t-1} z'_{t-1} + \ldots + z_1 z'_1)$$

so both S_t and S_t^{-1} are bounded.

The partial derivative of Q w.r.t. θ is only dependent on $\{z_t\}, \{S_t^{-1}\}$, so from the boundedness of W_t , we know that this algorithm satisfies (A.3).

(A.3) The limit $\lim_{t\to\infty} EQ(t, \theta, X_t) = h(\theta)$ exists for $\theta \in D_R$

We fix the value of θ in $Q(t, \theta_{t-1}, X_t)$ and compute the expectation over X_t. Fixing the value of θ means fixing the values of ϕ and S, so that we have

$$h_\phi(\phi, S) = lim_{t\to\infty} ES^{-1} z_{t-1}(z'_{t-1}(T(\phi) - \phi) + \eta_t)$$

$$h_s(\phi, S) = lim_{t\to\infty} \frac{t}{t+1} E(z_{t-1} z'_{t-1} - S)$$

Since

$$Ez_t z_t' = Ez_{t-1} z_{t-1}' = \begin{pmatrix} 1 & 0 \\ 0 & * \end{pmatrix} \equiv M$$

$Ez_{t-1}\eta_t = 0$, and $\lim_{t\to\infty} t/(t+1) = 1$, we obtain

$$h_\phi(\phi, S) = S^{-1}M(T(\phi) - \phi)$$

$$h_s(\phi, S) = M - S$$

where S and M are bounded, $T(\phi) - \phi$ is a linear function of ϕ, so the limit exists as long as ϕ stays in a bounded area.

(A.4) Since

$$Q_\phi(t, \theta_{t-1}, X_t) = S^{-1} z_{t-1}(z_{t-1}'(T(\phi) - \phi) + \eta_t)$$
$$Q_s(t, \theta_{t-1}, X_t) = vec((\frac{t}{t+1})(z_t z_t' - S_{t-1}))$$

According to Law of Large Numbers

$$\lim_{t\to\infty} q_{t,\theta}^{\phi} = \lim_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t} Q_\phi(s, \theta, X_s) = \lim_{t\to\infty} EQ_\phi(s, \theta, X_s) = h_\phi(\theta)$$

Also, $Q_s(t, \theta_{t-1}, X_t)$ is a very smooth function of t, so

$$\lim_{t\to\infty} q_{t,\theta}^{s} = \lim_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t} Q_s(s, \theta, X_s) = \lim_{t\to\infty} EQ_s(s, \theta, X_s) = h_s(\theta)$$

Now we explicitly compute the associated ODE using equation (3.20) and determine its stability conditions.

The associated ODE:

$$\frac{d\phi}{d\tau} = S^{-1}M(T(\phi) - \phi) \tag{3.30}$$

$$\frac{dS}{d\tau} = M - S \tag{3.31}$$

The second set of equations is a globally stable system with $S \to M$ from any starting point. It follows that $S^{-1}M \to 1$ from any starting point, provided S is invertible along the path, and hence the stability of the differential equations (3.30)-(3.31) is determined entirely by the stability of the smaller dimension system

$$\frac{d\phi}{d\tau} = T(\phi) - \phi \tag{3.32}$$

Note that equation(3.32) is identical to the differential equation (3.7) which defines E-stability. We have already seen that $\overline{\phi'} \equiv (\overline{a}, \overline{b'})$ is stable under equation (3.7) provided $\alpha < 1$. We can write

$$T(\phi) - \phi = \begin{pmatrix} \mu \\ \delta \end{pmatrix} + (\alpha - 1)I\phi$$

Equation (3.32) is a linear differential equation, all of whose eigenvalues of the Jacobian Matrix are equal to $\alpha - 1$. $\overline{\phi}$ is thus a globally stable equilibrium point of equation (3.32) if $\alpha < 1$, but is unstable if $\alpha > 1$. We also need to verify the boundedness of θ_t to apply our main theorem.

Write $T(\phi) = \alpha\phi + c$, with $0 < \alpha < 1$, so we have

$$\phi_t = \phi_{t-1} + (1/t)S_{t-1}^{-1} z_{t-1} z_{t-1}' ((\alpha - 1)\phi_{t-1} + c) + (1/t)S^{-1} z_{t-1} \eta_{t-1}.$$

We now restrict to the case $\eta = 0$.

Let $M_t = S_t^{-1} z_t z_t'$ and $f_t = \phi_t + (\frac{1}{1-\alpha})c$.

Then we have

$$f_t = (I - \frac{1-\alpha}{t} M_t) f_{t-1}.$$

Now M_t is positive semi-definite, so for t sufficiently large, the eigenvalues of the symmetric matrix $N_t = I - \frac{1-\alpha}{t} M_t$ lie in the interval $[0, 1]$.

So $|f_t| \leq ||N_t|| |f_{t-1}| \leq |f_{t-1}|$ and the sequence $\{f_t\}$, hence $\{\phi_t\}$, is bounded. S_t is bounded according to the argument above in (A.2).

Applying Theorem 3.4, it follows that under SRA (3.28)-(3.29), $(\phi_t, S_t) \to (\overline{\phi}, M)$ with probability 1, from any starting point, if $\alpha < 1$.

We recall that $\alpha = -r_p/m_p$ in the Cobweb model. So if we set p_t^e in the supply equation as p_t and we omit the noisy terms in both demand and supply equations in our model, α would simply be the slope ratio between demand curve and supply curve in a price-quantity coordinate. So we could intuitively explain our results in the following way:

If the slope of the supply curve is greater than the slope of the demand curve (in absolute value), then the fluctuations decrease in magnitude with each cycle, so a plot of the prices and quantities over time would look like an inward spiral, as shown in the diagram. This is called the stable or convergent case.

Otherwise, if the slope of the supply curve is less than the slope of the demand curve (in absolute value), then the fluctuations increase in magnitude with each cycle, so that prices and quantities spiral outwards. This is called the unstable or divergent case. The following graph shows that in the stable case, how initial price converges to the equilibrium price.

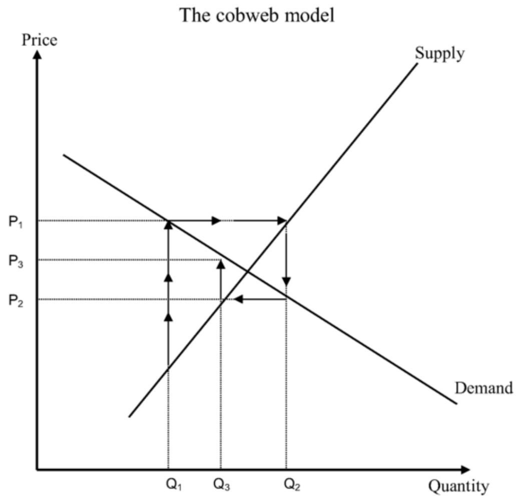

Figure 3.1: Cobweb Model with a stable equilibrium price

3.7 The E-stability Principle

When we consider the REE solutions of economic models, we assume that any particular solution of the model can be described as a stochastic process with particular parameter values $\overline{\phi}$. We will continue to find that the stability of $\overline{\phi}$ under learning can be determined by the E-stability equation

$$\frac{d\phi}{d\tau} = T(\phi) - \phi \tag{3.33}$$

in a neighborhood of $\overline{\phi}$, where $T(\phi)$ is a mapping from the perceived law of motion ϕ to the implied actual law of motion $T(\phi)$. Formally, $\overline{\phi}$ is said to be E-stable if it is locally asymptotically stable under the differential equation.

The correspondence between E-stability of an REE and its stability under adaptive learning we call the *E-stability Principle*. In our discussion of the Cobweb model above, we implicitly assumed that the agents knew the correct asymptotic specification. However, it is reasonable to ask how misspecification would alter the results. Suppose agents overparameterize the solution, e.g. suppose the REE being examined follows an ARMA process and the agents fit a process with higher AR or MA degree. Or suppose the REE being examined is a k-cycle and agents overfit with nk-cycle. Will such overfitting alter the stability conditions? It turns out that this is an important issue to be examined particularly when there are multiple REEs.

The issue of overparameterization has a simple reflection in terms of E-stability. If agents overparameterize an REE solution, the solution can be presented as a higher-dimensional vector $\widetilde{\phi}' = (\phi_1, \phi_2)$ with component values$(\overline{\phi}, 0)$ at the REE in question. We can now look at the stability of

$$\frac{d\widetilde{\phi}}{d\tau} = T(\widetilde{\phi}) - \widetilde{\phi} \tag{3.34}$$

Indeed, it is useful to introduce some terminology to represent this distinction. If the REE $\overline{\phi}$ is locally stable under equation (3.33) but $(\overline{\phi}, 0)$ is not locally stable under equation (3.34), then we say that $\overline{\phi}$ is *weakly E-stable*, while if $(\overline{\phi}, 0)$ is also stable under equation(3.34), then we say $\overline{\phi}$ is *strongly E-stable.*

Weak and strong E-stability govern whether the corresponding adaptive learning rule are stable. By expanding the dimension of ϕ appropriately, one can also allow for heterogeneous expectations across agents and determine whether allowing for heterogeneity alters the stability conditions for convergence of adaptive learning. For more details, refer to Evans and Honkapohja [6], Chapter 3. Also, these two concepts prove their importance in topic about convergence of REE in a stochastic nonlinear model. (see Evan and Honkapohja[4])

Appendices

A Proof of Lemma 3.5

The rigorous proof of Lemma 3.5 is presented below:

The way we choose $\rho(\theta^*)$ is that

$$\rho(\theta^*) < \frac{\delta(\theta^*)}{8M}$$

where $\delta(\theta^*) = V'(\theta^*)h(\theta^*) > 0$, $M = \max\{|V'(\theta)|, |V''(\theta)|, \mathcal{K}\}$. $\mathcal{K}$ represents for the bound of the first order partial derivative of Q w.r.t θ, θ could run every possible value in $\overline{D}$. In our following discussion, we write $\rho = \rho(\theta^*), \delta = \delta(\theta^*)$ for simplicity.

Then choose

$$\Delta\tau < \Delta\tau_0 = \min\{\frac{\rho}{4|h(\theta^*)|}, \frac{1}{8\mathcal{K}}\} \tag{1}$$

$$\epsilon_0(\rho) < \min\{\frac{1}{2}, \frac{\rho\mathcal{K}}{|h(\theta^*)|}\} \tag{2}$$

Introduce $j = j_{t_k}$ such that

$$t_k \leq j \leq m_{t_k,\Delta\tau}$$

and

$$\theta_s \in \mathcal{B}(\theta^*, 2\rho), \quad s = t_k, t_k + 1, \ldots, j - 1$$

An expression for θ_j

$$\begin{aligned}
\theta_j &= \theta_{t_k} + \sum_{s=t_k+1}^{j} \frac{1}{s} Q(s, \theta_{s-1}, X_s) \\
&= \theta_{t_k} + \sum_{s=t_k+1}^{j} \frac{1}{s} Q(s, \theta^*, X_s) + \sum_{s=t_k+1}^{j} \frac{1}{s}[Q(s, \theta_{s-1}, X_s) - Q(s, \theta^*, X_s)]
\end{aligned}$$

We have

$$|Q(s, \theta_{s-1}, X_s) - Q(s, \theta^*, X_s)| \leq \mathcal{K}|\theta_{s-1} - \theta^*| \leq \mathcal{K} \max_{t_k \leq k \leq j-1} |\theta_k - \theta^*| \tag{3}$$

Further more, we could write

$$\begin{aligned}
\theta_j &= \theta_{t_k} + \sum_{s=t_k+1}^{j} \frac{1}{s} Q(s, \theta^*, X_s) + \sum_{s=t_k+1}^{j} \frac{1}{s}[Q(s, \theta_{s-1}, X_s) - Q(s, \theta^*, X_s)] \\
&= \theta_{t_k} + \sum_{s=t_k+1}^{j} \frac{1}{s} h(\theta^*) + \sum_{s=t_k+1}^{j} \frac{1}{s}[Q(s, \theta^*, X_s) - h(\theta^*)] + \sum_{s=t_k+1}^{j} \frac{1}{s}[Q(s, \theta_{s-1}, X_s) - Q(s, \theta^*, X_s)]
\end{aligned}$$

or

$$\theta_j = \theta_{t_k} + \sum_{s=t_k+1}^{j} \frac{1}{s} h(\theta^*) + R^1_{t_k,j} + R^0_{t_k,j} \tag{4}$$

where

$$R^1_{t_k,j} = \sum_{s=t_k+1}^{j} \frac{1}{s}[Q(s,\theta^*,X_s) - h(\theta^*)]$$

and

$$|R^0_{t_k,j}| \le \sum_{s=t_k+1}^{j} \frac{1}{s}\mathcal{K} \max_{t_k \le k \le j-1} |\theta_k - \theta^*| \le 2\rho\Delta\tau\mathcal{K}(1+\epsilon_0) < 3\rho\Delta\tau\mathcal{K}$$

We want to show that

$$\max_{t_k \le j \le m_{t_k,\Delta\tau}} |R^1_{t_k,j}| \to 0 \text{ as } k \to \infty$$

so we introduce

$$q_{t_k,\theta^*} = \frac{1}{t_k} \sum_{s=1}^{t_k} Q(s,\theta^*,X_s)$$

Let the maximum of $R^1_{t_k,j}$ be attained for $j^* = j^*_{t_k}$, we have

$$q_{j^*,\theta^*} = q_{t_k,\theta^*} + \sum_{s=t_k+1}^{j^*} \frac{1}{s}[Q(s,\theta^*,X_s) - q_{s-1,\theta^*}] \tag{5}$$

By (A.4) we have

$$\lim_{k\to\infty} q_{t_k,\theta^*} = \lim_{k\to\infty} q_{j^*_{t_k},\theta^*} = \lim_{t\to\infty} EQ(t,\theta^*,X_t) = h(\theta^*)$$

Remark: the equation above holds naturally when $Q(t,\theta^*,X_t)$ is independent of t by the law of large numbers (Or a very smooth function of t). For more general cases, it requires more work to prove. Then

$$\begin{aligned}
\lim_{k\to\infty} R^1_{t_k,j^*} &= \lim_{k\to\infty} \sum_{s=t_k+1}^{j^*} \frac{1}{s}[Q(s,\theta^*,X_s) - h(\theta^*)] \\
&= \lim_{k\to\infty} \sum_{s=t_k+1}^{j^*} \frac{1}{s}[Q(s,\theta^*,X_s) - q_{s-1,\theta^*}] \\
&= \lim_{k\to\infty} (q_{j^*_{t_k},\theta^*} - q_{t_k,\theta^*}) = 0
\end{aligned}$$

so

$$\theta_j = \theta_{t_k} + \Delta\tau h(\theta^*) + R^1_{t_k,j} + R^2_{t_k,j} \quad \text{when} \quad j = m_{t_k,\Delta\tau}$$

where

$$R^1_{t_k,j} \to 0 \quad \text{when} \quad k \to \infty$$

and

$$|R^2_{t_k,j}| \leq |R^0_{t_k,j}| + |\epsilon_0 \Delta\tau h(\theta^*)| < 4\rho\Delta\tau\mathcal{K}$$

Now we want to show that $\theta_{m_{t_k,\Delta\tau}} \in \mathcal{B}(\theta^*, 2\rho)$

We shall show that

$$\theta_s \in \mathcal{B}(\theta^*, 2\rho), s = t_k, t_k + 1, \ldots, j - 1$$

imply that

$$\theta_j \in \mathcal{B}(\theta^*, 2\rho)$$

for sufficiently large k and small $\Delta\tau$. Then by induction it follows that

$$\theta_{m_{t_k,\Delta\tau}} \in \mathcal{B}(\theta^*, 2\rho)$$

We have

$$\begin{aligned} |\theta_j - \theta^*| &\leq |\theta_j - \theta_{t_k}| + |\theta_{t_k} - \theta^*| & (6) \\ &\leq |\sum_{s=t_k+1}^{j} \frac{1}{s} h(\theta^*)| + |R^1_{t_k,j}| + |R^2_{t_k,j}| + \rho & (7) \\ &\leq |\Delta\tau h(\theta^*)| + |\max_{t_k \leq j \leq m_{t_k,\Delta\tau}} |R^1_{t_k,j}| + 4\rho\Delta\tau\mathcal{K} + \rho & (8) \end{aligned}$$

Now choose k large so that

$$\max_{t_k \leq j \leq m_{t_k,\Delta\tau}} |R^1_{t_k,j}| < \rho/4 \qquad (9)$$

It now follows from (1), (10) and (11) that

$$|\theta_j - \theta^*| < |\Delta\tau h(\theta^*)| + \rho/4 + \rho/2 + \rho < 2\rho$$

By induction it follows that

$$\theta_{m_{t_k,\Delta\tau}} \in \mathcal{B}(\theta^*, 2\rho)$$

End proof of Lemma 3.5

B Proof of Theorem 3.4

The rigorous proof of Theorem 3.4 is presented below:

From assumption that $\overline{D}$ is a compact subset of $\mathcal{R}^n$, we know that for any sequence in $\overline{D}$, there exists a convergent subsequence. So for $\theta_t \in \overline{D}$, there exists a convergent subsequence θ_{t_k}. Let θ^* be the cluster point such that

$$\theta_{t_k} \to \theta^* \quad \text{as} \quad k \to \infty$$

where $\theta^* \neq \overline{\theta}$. Consequently, for arbitrarily small $\epsilon > 0$

$$|\theta_{t_k} - \theta^*| < \epsilon \qquad \text{for} \;\; k > K(\epsilon)$$

Consider now

$$V(\theta_{m_{t_k,\Delta\tau}}) - V(\theta_{t_k})$$

where $m_{t_k,\Delta\tau}$ is defined in Lemma 3.5. Denote $t_k = k'$ and $m_{t_k,\Delta\tau} = k''$, and use the Mean Value Theorem and Taylor expansion. This gives

$$V(\theta_{k''}) - V(\theta_{k'}) = V'(\xi)[\theta_{k''} - \theta_{k'}] = V'(\theta^*)[\theta_{k''} - \theta_{k'}] + (\xi - \theta^*)^{\tau} \cdot V''(\xi')[\theta_{k''} - \theta_{k'}] \quad (1)$$

where ξ and ξ' belong to $\mathcal{B}(\theta^*, 2\epsilon)$.

Now take $\epsilon < \Delta\tau\rho$, where $\rho < \frac{\delta(\theta^*)}{8M}$, and apply Lemma 3.5 to θ_{t_k}, we have

$$\theta_{k''} - \theta_{k'} = \Delta\tau h(\theta^*) + q^1_{k',k'',\theta^*} + q^2_{k',k'',\theta^*} \quad (2)$$

where q^i are subject to (3.26),(3.27).

Insert (2) into (1)

$$V(\theta_{k''}) - V(\theta_{k'}) = \Delta\tau V'(\theta^*)h(\theta^*) + R_{\Delta\tau,t_k,\theta^*} \quad (3)$$

where

$$R_{\Delta\tau,t_k,\theta^*} = V'(\theta^*)[q^1_{k',k'',\theta^*} + q^2_{k',k'',\theta^*}] + (\xi - \theta^*)^{\tau} \cdot V''(\xi')[\theta_{k''} - \theta_{k'}]$$

Apply Taylor expansion again

$$V(\theta_{k''}) = V(\theta^*) + (\theta_{k'} - \theta^*)V'(\xi''') + \Delta\tau V'(\theta^*)h(\theta^*) + R_{\Delta\tau,t_k,\theta^*} \quad (4)$$

where $\xi''' \in \mathcal{B}(\theta^*, 2\epsilon)$

Now since that the cluster point θ^* is different from the equilibrium point $\overline{\theta}$. Then $V'(\theta^*)h(\theta^*) = -\delta, \delta > 0$. So we have

$$V(\theta_{k''}) = V(\theta^*) - \Delta\tau\delta + (\theta_{k'} - \theta^*)V'(\xi''') + R_{\Delta\tau,t_k,\theta^*} \quad (5)$$

Now we take care of the part

$$(\theta_{k'} - \theta^*)V'(\xi''') + R_{\Delta\tau,t_k,\theta^*}$$

where $|\theta_{k'} - \theta^*| < \epsilon$, we know that

$$|R_{\Delta\tau,t_k,\theta^*}| = |V'(\theta^*)[q^1_{k',k'',\theta^*} + q^2_{k',k'',\theta^*}] + (\xi - \theta^*)^{\tau} \cdot V''(\xi')[\theta_{k''} - \theta_{k'}]|$$

where $|\theta_{k''} - \theta_{k'}| < \epsilon$, $q^1_{k',k'',\theta^*} \to 0$, $|q^2_{k',k'',\theta^*}| \leq 4\Delta\tau\mathcal{K}\rho$

We choose $k > K'$ big enough so that $|q^1_{k',k'',\theta^*}| < \epsilon$, then by the choose of ρ, ϵ, k', we get

$$|(\theta_{k'} - \theta^*)V'(\xi''') + R_{\Delta\tau,t_k,\theta^*}| < \epsilon \cdot M + \epsilon \cdot M + \epsilon^2 \cdot M + 4\Delta\tau\rho \cdot M$$

Since we know $\epsilon < \Delta\tau\rho$, $\rho < \delta/8M$, we could write

$$(\theta_{k'} - \theta^*)V'(\xi''') + R_{\Delta\tau,t_k,\theta^*} < 3\delta\Delta\tau/8 + \delta\Delta\tau/2 = \frac{7\delta\Delta\tau}{8}$$

so

$$V(\theta_{m_{t_k,\Delta\tau}}) < V(\theta^*) - \Delta\tau\delta/8 \tag{6}$$

where $k > \max\{K(\epsilon), K'\}$

With (6) holds for any subsequence θ_{t_k} that converges to a point different from $\overline{\theta}$, we conclude that

$$\lim_{t\to\infty} \inf V(\theta_t) = 0$$

Consider $\inf V(\theta)$ taken over all cluster points in D_A. Let this value be U. Since the set of cluster points in D_A is compact, there exists a cluster point $\hat{\theta}$, such that $V(\hat{\theta}) = U$. If now $U > 0$, $V'(\hat{\theta})h(\hat{\theta})$ will be strictly negative ($= -\delta'$) and from (6), there exists $V(\theta_k*)$ takes a value less than $U - \hat{\delta}\Delta\tau/8$ infinitely often, which contradicts U being the infinum. Hence $U = 0$, which means that $\overline{\theta}$ is a cluster point.

We will show that

$$\lim_{t\to\infty} \sup V(\theta_t) = 0 \tag{7}$$

Suppose that

$$\lim_{t\to\infty} \sup V(\theta_t) = \overline{W} > 0$$

Since the set of cluster points in D_A is compact, there exists a cluster point $\widetilde{\theta}$, such that $V(\widetilde{\theta}) = W$.

Consider the interval $I = [W/3, 2W/3]$. Since $\overline{\theta}$ is a cluster point and since $V(\theta_t)$ is supposed to have a subsequence tending to W, this interval I is crossed ‘upwards’ and ’downwards’ infinitely many times by $V(\theta_t)$.

Since

$$|\theta_t - \theta_{t-1}| = |\frac{1}{t}Q(t, \theta_{t-1}, X_t)| \leq |\frac{1}{t}Q(t, \overline{\theta}, X_t)| + |\frac{1}{t}|\mathcal{K}|\theta_{t-1} - \overline{\theta}|$$

According to the proof of Lemma, $\frac{1}{t}Q(t, \overline{\theta}, X_t)$ goes to $\frac{1}{t}h(\theta^*)$ when t goes very large. Also, the second term tends to zero. So the step size $\theta_t - \theta_{t-1}$ tends to zero, and hence there will be a subsequence of $V(\theta_t)$ entirely in the interval I. Consider now a special sequence of "upcrossings":

Let the sequence t'_k be defined such that

$$V(\theta_{t'_k - 1}) < W/3 \tag{8}$$

$$V(\theta_{t'_k}) \geq W/3 \tag{9}$$

$$V(\theta_{t'_k + s_k}) > 2W/3 \tag{10}$$

where

$$s_k \text{ is the first } s \text{ for which } \; V(\theta_{t_k + s}) \text{ not in } I \tag{11}$$

$$\theta_{t'_k} \to \theta^{**} \quad \text{ as } \; k \to \infty$$

It is clear that $V(\theta^{**}) = W/3$ and let $V'(\theta^{**})h(\theta^{**}) = -\delta^{**}$. From (6) we now have that

$$V(\theta_{m_{t'_k, \Delta\tau}}) < W/3 - \delta^{**}\Delta\tau/8 \tag{12}$$

This means that $V(\theta_{t'_k + s'_k})$ is not in I where $s'_k = m_{t'_k, \Delta\tau} - t'_k$.

Suppose now that there is a $s_k < s'_k$ such that $V(\theta_{t'_k + s_k}) > 2W/3$. However, V is continuous and $\theta_{t'_k + s_k}(s_k < m_{t'_k, \Delta\tau} - t'_k)$ will belong to an arbitrarily small neighborhood of $\theta_{t'_k}$ for sufficiently small $\Delta\tau$ according to Lemma 3.5. Therefore (12) implies a contradiction to the existence of a subsequence t'_k with properties (8)-(11). Now another case is none of these $s_k < s'_k$ satisfies that $V(\theta_{t'_k + s_k}) > 2W/3$. Thus, all terms between $\theta_{t'_k}$ and $\theta_{m_{t'_k, \Delta\tau}}$ do not have a V value which exceeds $2W/3$. Also, $V(\theta_{m_{t'_k, \Delta\tau}}) < W/3$. Consider the next step for θ_t which enters interval I, we could repeat the argument above, so all terms after $\theta_{t'_k}$ stay less than $2W/3$. Hence, no interval I may exist. W must be zero and (7) follows. This concludes $\theta_t \to \overline{\theta}$.

References

[1] Andrew, J.Culham. (2007) Asset Pricing in a Lucas Framework with Boundedly Rational, Heterogeneous Agents. *Dissertation, the Florida State University, Department of Mathematics*

[2] Bray, M.M. and Savin, N. E.(1986) Rational Expectations Equilibria, Learning, and Model Specification. *Econometrica,* Vol.54, No.5, pp.1129-1160

[3] Cramér, H., and Leadbetter M.R.(1967) *Stationary and Related Stochastic Processes* Newyork, Wiley

[4] Evans, G.W. and Honkapohja, S. (1995) Local Convergence of Recursive Learning to Steady States and Cycles in Stochastic Nonlinear Models.*Econometrica*, Vol.63, No.1, pp195-206

[5] Evans, G.W. and Honkapohja, S. (1998) Economic Dynamics with Learning: New Stability Results. *Review of Economic Studies*, 65, pp.23-44

[6] Evans, G.W. and Honkapohja, S. (2001) *Learning and Expectations in Macroeconomics.* Princeton University Press

[7] Hirsch, M.W. amd S. Smale (1974) *Differential Equations, Dynamic Systems, and Linear Algebra.* Academic Press,Orlando,FL

[8] Krasovskii, N. (1963) *Stability of Motion.* Stanford University Press

[9] Ljung,L. (1977) Analysis of Recursive Stochastic Algorithms. *IEEE Thansactions on Automatic Control*,22, pp551-575

[10] Ljungqvist, L. and Sargent, T.J. (2000) *Recursive Macroeconomic Theory.* MIT Press

[11] Woodward,M. (1990) Learning to Believe in Sunspots. *Econometrica*, Vol.58, pp277-307