Center for Economic Studies, University of Munich, 80539 Munich, Germany

CENTER FOR ECONOMIC STUDIES

ECONOMIC DYNAMICS WITH LEARNING: NEW STABILITY RESULTS

Seppo Honkapohja
George W. Evans

Working Paper No. 109

UNIVERSITY OF MUNICH

CES

Working Paper Series

CES Working Paper Series

ECONOMIC DYNAMICS WITH LEARNING: NEW STABILITY RESULTS

Seppo Honkapohja
George W. Evans

Working Paper No. 109

1996

Center for Economic Studies
University of Munich
Schackstr. 4
80539 Munich
Germany
Telephone & Telefax:
++49-89-2180-3112

Support from *CES* at Munich, SPES Program of the EU, and the grants to RUESG at University of Helsinki is gratefully acknowledged.

CES Working Paper No. 109
May 1996

ECONOMIC DYNAMICS WITH LEARNING: NEW STABILITY RESULTS

Abstract

Drawing upon recent contributions in the statistical literature, we present new results on the convergence of recursive, stochastic algorithms which can be applied to economic models with learning and which generalize previous results. The formal results provide probability bounds for convergence which can be used to describe the local stability under learning of rational expectations equilibria in stochastic models. Economic examples include local stability in a multivariate linear model with multiple equilibria and global convergence in a model with a unique equilibrium.

Seppo Honkapohja
Department of Economics
University of Helsinki
P.O. Box 54
00014 Helsinki
Finland

George W. Evans
Department of Economics
University of Oregon
Eugene, OR 97403-1285
USA

ECONOMIC DYNAMICS WITH LEARNING: NEW STABILITY RESULTS

by George W. Evans and Seppo Honkapohja

1. Introduction

Adaptive learning in dynamic expectations models has been the subject of extensive research effort in the recent economics literature (for a survey of the topic see Evans and Honkapohja (1995b) or Sargent (1993)). Much of this research has been devoted to obtaining stability conditions for convergence of learning dynamics to a rational expectations equilibrium (REE). For example, Bray and Savin (1986) and Fourgeaud, Gourieroux and Pradel (1986) provide conditions under which the unique REE is globally stable under learning. Woodford (1990) shows that for some specifications of the standard overlapping generations model there will be convergence to the set of stationary sunspot equilibria.

Marcet and Sargent (1989a) have shown that convergence results can be obtained for general linear models, having a unique REE, using the stochastic approximation technique of Ljung (1977). When a model has multiple REE it is of interest to analyze the local stability properties of any given REE, and Evans and Honkapohja (1994a,b, 1995a,b) have shown how these techniques can be used in a variety of such models.

Most of the convergence results of Marcet and Sargent (1989a,b) and Evans and Honkapohja (1994a,b, 1995a,b) are based on Theorem 4 of Ljung (1977) which employs a "projection facility" constraining estimates to remain in a region around the REE. Grandmont and Laroque (1991) and Grandmont (1994) suggest that this is vital for the stability results but argue that the assumption of a projection facility is inappropriate for decentralized markets.[1] This raises the issue of the precise role of the projection facility and whether useful general results can be obtained when it is not employed.

To motivate the paper more concretely, we introduce a simple example which we will

----

[1]Moreno and Walker (1994) also stress difficulties arising from the use of a projection facility. Benassy and Blad (1989) discuss related instability results.

consider at some length later. Suppose that the value of an economic variable of interest y_t depends on its lagged value, its future expected value and a white noise shock v_t, according to the reduced form model

$$y_t = \beta y^e_{t+1} + \delta y_{t-1} + v_t. \tag{1.1}$$

Here y^e_{t+1} denotes the expectation of y_{t+1} based on information available at time t, which is assumed to include current and past values of y_t. Provided the roots of $\beta b^2 - b + \delta = 0$ are real, there are two REE solutions of the form

$$y_t = b^* y_{t-1} + d^* v_t, \tag{1.2}$$

where $d^* = (1-\beta b^*)^{-1}$. Consider a root b^* with $|b^*| < 1$, so that the corresponding REE (1.2) is asymptotically stationary, and ask whether agents who are not initially endowed with rational expectations would be able to learn b^* using a statistical learning rule such as recursive least squares.

Thus, following the earlier literature cited above, we assume that agents believe that data is being generated by a law of motion of the form (1.2), but that b^* is unknown to them and estimated at each time t from available data using the recursive least squares rule[2]

$$b_t = b_{t-1} + t^{-1} R_t^{-1} y_{t-2}(y_{t-1} - b_{t-1} y_{t-2})$$
$$R_t = R_{t-1} + t^{-1} (y^2_{t-2} - R_{t-1}).$$

Here b_t is the least squares estimate of b^* and R_t is the estimate of the second moment of y_t. It can be readily verified that the standard formula for a least squares estimate of b_t based on a simple regression of y_t on y_{t-1} satisfies these recursive equations.

Under these assumptions the complete dynamic system is as follows: given b_t, agents

----

[2]We simplify here by assuming that b_t is constructed using data only through period t-1.

make forecasts $y^e_{t+1} = b_t y_t$ and y_t is generated by these forecasts and by the exogenous shock v_t according to the model (1.1). The new data point y_t is then used the following period to revise the estimate of b^* to b_{t+1} according to the recursive least squares formula. Continuing in this way, the system evolves over time.

The central question of interest is whether it can be established that b_t converges to b^* as $t \to \infty$, so that adaptive agents, following a standard statistical learning rule, eventually learn to have rational expectations. Following the procedures of Marcet and Sargent (1989a) and Evans and Honkapohja (1994a) the following results can be shown (generically). If an appropriate local stability condition at b^* is met, *and provided the recursive least squares algorithm is modified to include a suitable Projection Facility*, then $b_t \to b^*$ with probability 1.[3] If instead the stability condition is not met then $b_t \to b^*$ with probability 0. It is on this basis that it has been proposed that the local stability condition (often called the "expectational stability" or E-stability" condition) be used to classify REE solutions as locally stable or unstable under adaptive learning.

These are very useful results, but the convergence results are obtained at the cost of a possibly very strong assumption, namely the incorporation of a projection facility. Briefly (and informally) the projection facility is a technical device which constrains estimates b_t never to leave some prespecified neighborhood of b^* even if they would do so under the unmodified recursive least squares algorithm.[4] To obtain the convergence result the projection facility may need to be a "small" neighborhood of b^* in order to prevent estimates from straying into regions where they would diverge. It is clear that the projection facility is in general essential to obtaining probability 1 convergence results in stochastic models with multiple REE[5]. The issue opened by the criticisms of Grandmont and

----

[3]Also R_t converges to Ey_t^2 as b_t converges to b^*.

[4]One version of the projection facility would reset b_t back to some previous value if it would otherwise leave the specified neighborhood under the algorithm. See Marcet and Sargent (1989a) for a full discussion.

[5]In models with unique REE, e.g. the "cobweb" model studied by Bray and Savin (1986), it is

Laroque is how critical the projection facility is to obtaining positive results of *any* kind in such models.

For a nonstochastic model of the above form Grandmont and Laroque show that there is an open set of initial conditions near the nonstochastic steady state which lead to local divergence even when the E-stability condition is met at that solution, *unless* a sufficiently small projection facility is in place. This clearly raises doubts about the classification criterion above which has been proposed for stochastic versions of the model, since these rely on the projection facility. Might it even be the case that an REE to which the system converges almost surely *with* a projection facility in place becomes wholly unstable when a projection facility is *not* employed? Can robust local convergence results be stated *without* a projection facility which might justify the classification scheme in models with multiple REE?

These are the main issues addressed in this paper. We show that recent results in the theory of stochastic approximation can be invoked to shed light on the questions. Using these techniques we show that, in stochastic models, convergence to an equilibrium satisfying a stability condition does obtain generally, though of course with probability less than one. The principal contribution of this paper is to state and apply these results and demonstrate how precise positive convergence results can be obtained, even in models with multiple equilibria, without resorting to the controversial projection facility.

The main text of this paper describes these principal results in the context of the conditionally linear state variable dynamics which arise in most of the economics literature. The assumptions in this case are straightforward to interpret, and our main focus is to show how they can be applied in dynamic linear expectations models. A companion paper, Evans and Honkapohja (1996) provides the more general technical results for nonlinear Markovian setups. A statement of these results is provided in Appendix A-1, where we

----

possible to show global probability 1 convergence without a projection facility. However, because it affords technical simplifications, the projection facility has even been invoked in models with a unique REE. For further discussion see sections 4 and 5 below.

demonstrate that the general Markovian framework covers the conditional linear dynamics set-up applied in this paper.

A second objective of the paper is to provide both local and global results on the convergence of least squares learning in general multivariate linear economic models. Such models occur frequently in macroeconomics but they have not been previously examined in the learning literature. These applications also demonstrate that the framework and the basic convergence results of this paper can be applied to economic models.

Our first application is a multivariate linear model, with multiple autoregressive REE. This generalizes the above example by permitting y_t to be a vector and by including a vector of exogenous observables which follow a vector autoregresssion. We show how to identify the condition for an equilibrium to be locally stable under learning.

The second example is a multivariate linear model with a unique rational expectations solution. The global convergence result for this model generalizes the corresponding results of Bray and Savin (1986) and Marcet and Sargent (1989a). Heretofore these generalizations have also only been achieved using (arbitrarily large) projection facilities.

2. The Algorithms and Basic Results.

The basic convergence results are described in this section in the context of an abstract recursive stochastic algorithm. Subsequent sections show how this framework can be applied to standard dynamic macroeconomic expectations models. Our results are based on the analysis of stochastic approximation techniques developed in Benveniste, Métivier and Priouret (1990).

Consider the algorithm

$$\theta_t = \theta_{t-1} + \gamma_t H(\theta_{t-1},X_t) + \gamma_t^2 \rho_t(\theta_{t-1},X_t). \tag{2.1}$$

Here the vector of parameter estimates θ_t lies in $\mathbb{R}^d$ and the observable state vector X_t lies in $\mathbb{R}^k$. The assumptions on the stochastic process for X_t are given as (B.1)-(B.3) below.

Intuitively, θ_t is a vector of parameters which are being recursively updated, e.g.

the parameters of an agent's forecasting rule. $H(\theta_{t-1},X_t)$ is a time-homogeneous function which states how θ_t is to be updated as a result of the most recent observation of the system, and γ_t is a sequence of scalar "gain" parameters which specifies the size of the response of θ_t to H at time t. The final term $\rho_t(\theta_{t-1},X_t)$ allows for a second-order time-varying dependence on the system (this term is sometimes not present). In the context of the simple example in the introduction, $\theta_t = (b_t,R_t)$, $X_t = (y_{t-1}, y_{t-2})$, and $\gamma_t = t^{-1}$.

Fix an open subset $D \subseteq \mathbb{R}^d$. We make the following assumptions:

(A.1) γ_t is a nonstochastic nonincreasing sequence satisfying

$$\sum \gamma_t = +\infty \text{ and } \sum \gamma_t^2 < +\infty.$$

(A.2) For any compact subset $Q \subset D$, there exists C_1, C_2, q_1 and q_2 such that $\forall\theta\in Q$ $\forall t$:

(i) $|H(\theta,x)| \le C_1(1+|x|^{q_1})$,

(ii) $|\rho_t(\theta,x)| \le C_2(1+|x|^{q_2})$.

We remark that the assumption $\sum \gamma_t = +\infty$ is required to avoid convergence of θ_t to a nonequilibrium point and the assumption $\sum \gamma_t^2 < +\infty$ guarantees elimination of residual fluctuation in θ_t asymptotically. Assumption (A.1) is stronger than necessary. In fact, $\sum \gamma_t = +\infty$ plus the condition $\sum \gamma_t^p < +\infty$ for some $p \ge 2$ is sufficient. However, most economic applications in the literature satisfy (A.1). (A.1) is satisfied by $\gamma_t = t^{-1}$, the standard assumption for least squares learning. (A.2) simply says that H and ρ_t are polynomial bounded in the state variable.

Next, we assume a set of Lipschitz conditions:

(A.3) For any compact subset $Q \subset D$ the function $H(\theta,x)$ satisfies for all $\theta,\theta' \in Q$, and $x_1,x_2,x \in \mathbb{R}^k$:

(i) $|H(\theta,x_1)-H(\theta,x_2)| \le L_1|x_1-x_2|$,

(ii) $|H(\theta,0)-H(\theta',0)| \le L_2|\theta-\theta'|$,

(iii) $|\partial H(\theta,x)/\partial x-\partial H(\theta',x)/\partial x| \le L_2|\theta-\theta'|$,

for some constants L_1, L_2.

Clearly (A.3) is satisfied if $H(\theta,x)$ is twice continuously differentiable with bounded second derivatives on every compact $Q \subset D$. We will see below that these assumptions are straightforward to verify for models such as the one presented in the Introduction above.

We also need assumptions on the stochastic process of the state variable $X_t \in \mathbb{R}^k$. In Appendix 1 we discuss the required assumptions for the convergence result for the general nonlinear case, where X_t is a Markov process with a transition probability law $\Pi_\theta(x,A)$ which may depend on θ_{t-1}. This result is based on certain assumptions which are not particularly intuitive.

In much of the economic literature it has turned out that X_t in fact follows *conditionally linear dynamics* (see Marcet and Sargent (1989a,b), Woodford (1990) and Evans and Honkapohja (1994b, 1995a,b) among others). For this case the requisite assumptions are relatively straightforward to state and verify. Consequently we postulate this case here:

(B.1) $X_t = A(\theta_{t-1})\, X_{t-1} + B(\theta_{t-1})\, W_{t-1}$.

(B.2) W_t is identically and independently distributed with finite absolute moments, i.e. $E|W_t|^q < \infty$ for all $q=1,2,3,\ldots$

(B.3) For any compact subset $Q \subset D$,

$$\sup_{\theta\in Q} |B(\theta)| \le M \text{ and } \sup_{\theta\in Q} |A(\theta)| \le \rho < 1,$$

for some matrix norm $|.|$, and $A(\theta)$ and $B(\theta)$ satisfy Lipschitz conditions on Q.[6]

We will see that it is also straightforward to verify these assumptions for models such as the one described in the introductory section.

Convergence of recursive algorithms such as (2.1) can be analyzed using an *associated differential equation*

$$d\theta/d\tau = h(\theta(\tau)). \qquad (2.2)$$

In the general Markovian case the function h(.) is given by

$$h(\theta) = \int H(\theta,y)\Gamma_\theta(dy),$$

where $\Gamma_\theta(.)$ is the unique invariant probability distribution of the transition probability law $\Pi_\theta(x,A)$ for each fixed θ. In the linear setup the function $h(\theta)$ can alternatively be obtained as follows[7]:

Lemma: For any θ the function

$$h(\theta) = \lim_{t\to\infty} E(H(\theta,\bar{X}_t(\theta)),$$

where

$$\bar{X}_t(\theta) = A(\theta)\,\bar{X}_{t-1}(\theta) + B(\theta)W_{t-1},$$

is well-defined and $h(\theta)$ is locally Lipschitz.

Theorem 1 below gives bounds on the probability of convergence or divergence of the algorithm. It will be seen that convergence is governed by the properties of $d\theta/d\tau=h(\theta(\tau))$. Intuitively, for values of t the trajectories θ_t of the algorithm are approximated by the time paths of (2.2) at specific points of time $\tau(t)$, where $\tau(t) = \sum_{i=0}^{t} \gamma_i$.

----

[6]The condition on $A(\theta)$ is somewhat stronger than having the spectral radius $r(A(\theta))<1$ for all $\theta\in Q$. However, note that if $r(A(\theta^*))<1$ at some θ^* then our condition holds in a neighborhood of θ^*.

[7] Proofs of the Lemma and all subsequent results are collected in Appendix 2.

Let $\theta^*\in D$ be an asymptotic stable equilibrium point of this differential equation. Theorem 1 is stated most naturally and generally using the Lyapunov contour sets of θ^*, which we now introduce (but see the Remark following the theorem). It is well known that there exists a C^2 Lyapunov function $U(\theta)$ on the domain of attraction $\mathcal{D}$ of θ^*. $U(\theta)$ satisfies:

(i) $U(\theta^*) = 0$, $U(\theta)>0$ for all $\theta\in\mathcal{D}$, $\theta\neq\theta^*$.

(ii) $U'(\theta)h(\theta)<0$ for all $\theta\in\mathcal{D}$, $\theta\neq\theta^*$.

(iii) $U(\theta) \to \infty$ if $\theta \to \partial\mathcal{D}$ or $|\theta| \to +\infty$,

where $\partial\mathcal{D}$ denotes the boundary of $\mathcal{D}$.

Consider the compact sets defined by the contours of $U(\theta)$:

$$K(c) = \{\theta; U(\theta)\le c\},$$

for $c\ge 0$. We can now state the basic result which is:

Theorem 1: Let θ^* be an asymptotically stable equilibrium point of the differential equation $d\theta/d\tau=h(\theta(\tau))$. Suppose that assumptions A and B are satisfied on $D=\mathrm{int}(K(c))$ for some $c>0$. Let $P_{n,x,a}$ be the probability distribution of $(X_t,\theta_t)_{t\ge n}$ with $X_n=x$ and $\theta_n=a$. Then for any compact $Q\subset D$ there exist constants F and s, depending on Q but not on $\{\gamma_t\}$, such that $\forall n\ge 0$, $a\in Q$, x:

$$P_{n,x,a}\{\theta_t \to \theta^*\} \ge 1 - F(1+|x|^s)J(n),$$

where J(n) is a positive decreasing sequence with $\lim_{n\to\infty} J(n)=0$. J(n) is in fact given by

$$J(n) = (1 + \sum_{k=n+1}^{\infty} \gamma_k^2)(\sum_{k=n+1}^{\infty} \gamma_k^2).$$

Remark: Suppose assumptions A and B are satisfied on some compact set $N\subset\mathcal{D}$ with θ^* in its interior. Then there exists an open ball $D\subset N$ around θ^* on which Theorem 1 holds.

The central content of Theorem 1 is that for stable equilibria it is possible to provide a lower bound for the probability of convergence. Moreover, the bound approaches one as time goes to infinity. That is, fix the compact neighborhood Q and suppose θ_n lies in Q at some time n where n is large. Then the probability that $\theta_t \to \theta^*$ is near 1.

Our treatment emphasizes the primitive assumptions required on the algorithm and state dynamics to obtain asymptotic convergence results. Theorem 1 and the Corollaries below can be readily used to analyze properties of learning algorithms in a wide range of economic models, as illustrated by our applications below. The framework here could also be used to analyze the models in Marcet and Sargent (1989a). We remark that, despite our dropping of the projection facility, we do not require their "difficult to verify" assumptions (A.6) and (A.7.1).

In the appendix we verify that assumptions B are a special case of assumptions C for the general Markovian case. A companion paper Evans and Honkapohja (1996) provides a proof, based on the analysis of Benveniste et. al. (1990), of Theorem 1 for the general Markovian case stated in Appendix 1.

The following two corollaries provide probability statements for the algorithm starting at the initial time 0. Using the expression for J(n) we immediately have from Theorem 1:

Corollary 1.1: Suppose $\gamma_t=\xi\gamma'_t$, where γ'_t satisfies (A.1). Consider initial values θ belonging to some compact domain Q in D. $\forall\delta>0$ $\exists\xi^*$ such that $\forall 0<\xi<\xi^*$ and $a\in Q$ we have

$$P_{0,x,a}\{\theta_t \to \theta^*\} \geq 1 - \delta.$$

Corollary 1.1 is the case of slow adaption. It shows that the probability of convergence can be made arbitrarily close to one, provided that the adaption rates are sufficiently low. For general adaption speeds and some auxiliary assumptions one can obtain convergence with positive probability for initial conditions sufficiently close to the

equilibrium:

Corollary 1.2: Assume that θ^* is locally asymptotically stable for $d\theta/d\tau=h(\theta(\tau))$. Assume also the existence of points x^* and w^* such that (θ^*,x^*) is invariant under (2.1) and (B.1), i.e.

$$H(\theta^*,x^*) = 0,\ \rho_t(\theta^*,x^*) = 0,$$
$$x^* = (I - A(\theta^*))^{-1}B(\theta^*)w^*,$$

and that for all $\eta>0$ $\mathrm{Prob}[|W_t-w^*|\leq\eta] > 0$. Then $\exists p>0$ such that for $|a-\theta^*|$ and $|x-x^*|$ sufficiently small we have

$$P_{0,x,a}\{\theta_t \to \theta^*\} \geq p.$$

It is not in general possible to obtain bounds close to one even for the most favorable initial conditions. The reason is that for small t the algorithm is not well approximated by the associated differential equation. Sufficiently large random shocks may displace θ_t outside the domain of attraction of the differential equation. (This problem does not arise in Corollary 1.1, because of slow adaption.)[8]

Finally, we remark that in the companion paper we show that with these assumptions one can derive as a corollary a version of Ljung's (1977) Theorem 4, which gives convergence with probability 1 when the algorithm is augmented to include a projection facility.

3. Applications to Linear Models

a. A multivariate model

We examine the multivariate model

----

[8]For certain setups it is possible to obtain convergence with probability one for initial values sufficiently near the equilibrium θ^*, provided the support of the distribution of the random shock W_t is small enough. See Evans and Honkapohja (1995a). We also remark that for particular models the neighborhood giving positive probability of convergence in Corollary 1.2 can in principle be "large" and need not even be confined to the basin of attraction of θ^* of the associated differential equation.

$$y_t = \alpha + \beta y^e_{t+1} + \delta y_{t-1} + \kappa w_t + v_t, \tag{3.1a}$$

$$w_t = \varphi w_{t-1} + e_t \tag{3.1b}$$

where y_t is an nx1 vector of endogenous variables, w_t is a px1 vector of observable exogenous variables and e_t is a vector of white noise shocks with bounded absolute moments. v_t is an nx1 white noise disturbance, with bounded absolute moments, assumed independent of w_t. It is assumed that the nxn matrix β is invertible and that the eigenvalues of the pxp matrix φ are inside the unit circle. δ is an nxn matrix and κ is nxp. Here y^e_{t+1} denotes the (possibly non-rational) expectations of y_{t+1} formed by agents at time t. This set-up corresponds to the system considered by McCallum (1983, Appendix).

This model has various REE. We focus here on solutions of the "minimal state variable" form (see McCallum 1983):

$$y_t = a + b\, y_{t-1} + c\, w_t + d\, v_t, \tag{3.2}$$

wher a, b, c, and d are conformable matrices. To determine the values of a, b, c and d we insert

$$E_t y_{t+1} = a + b\, y_t + c\varphi\, w_t$$

into (3.1a). Equating coefficients with (3.2) yields the matrix equations

$$\alpha = (I - \beta b - \beta)\, a \tag{3.3a}$$

$$b = \beta b b + \delta \tag{3.3b}$$

$$c = \beta b c + \beta c \varphi + \kappa. \tag{3.3c}$$

$$d = (I - \beta b)^{-1}. \tag{3.3d}$$

(3.3a), (3.3c) and (3.3d) have a unique solution (generically) given a solution b to (3.3b).[9] Generically, (3.3b) can have up to 2n distinct solutions for b (see Theorem 4 of Gantmacher (1959, Chapter 8)). McCallum focuses on a specific solution based on a subsidiary selection criterion. We do not restrict our attention to this choice since our interest is in providing conditions for the local stability of solutions of the form (3.2) under adaptive learning.

----

[9]For the formula of the solution for c, see Horn and Johnson (1991, p. 255).

We set up the learning scheme as follows. Define the state vector and parameter matrix as

$$z_t' = (1, y_{t-1}', w_t'),$$

$$\phi_t' = (a_t, b_t, c_t).$$

It is assumed that at time t agents forecast using the assumed law of motion $y_{t+1} = \phi_t' z_{t+1} + \eta_{t+1}$, where $E_t \eta_{t+1} = 0$. Inserting $y^e_{t+1} = a_t + b_t y_t + (c_t \varphi) w_t$ into (3.1a) we obtain the actual law of motion

$$y_t = T(\phi_t)' z_t + W(b_t) v_t, \tag{3.4}$$

where

$$T(\phi)' = [(I-\beta b)^{-1}(\alpha + \beta a),\ (I-\beta b)^{-1}\delta,\ (I-\beta b)^{-1}(\kappa + \beta c \varphi)]$$

and

$$W(b) = (I - \beta b)^{-1},$$

for $\phi' = (a,b,c)$, provided $I-\beta b$ is invertible. For the statement of the Proposition we also need to define $\mathcal{T}(\phi) = vec(T(\phi))$, where vec denotes the vectorization of a matrix. Let $D\mathcal{T}(\phi)$ denote its Jacobian with respect to $vec(\phi)$ at point ϕ.

In defining the algorithm to update ϕ_t we use a modification of recursive least squares. Let (ϕ_t, R_t) be defined by

$$\phi_t = \phi_{t-1} + \gamma_t R_t^{-1} z_{t-1}(y_{t-1} - \phi_{t-1}' z_{t-1})', \tag{3.5a}$$

$$R_t = R_{t-1} + \gamma_t (z_{t-1} z_{t-1}' - R_{t-1}). \tag{3.5b}$$

Note that we have assumed that the estimates ϕ_t are based on data only through period t-1, even though forecasts use information z_t. This is to avoid the complication of having z_t and ϕ_t simultaneously determined. For (3.5a-b) we make the additional technical assumption $\limsup_{t\to\infty} (1/\gamma_{t+1}) - (1/\gamma_t) < \infty$. It is easy to show that this holds for $\gamma_t = \xi t^{-1}$.

It may be noted that for $\gamma_t = 1/t$ and appropriate initial conditions the algorithm reduces to recursive least squares, i.e.

$$\phi_t = (\textstyle\sum_{j=1}^{t-1} z_j z_j')^{-1} (\sum_{j=1}^{t-1} z_j y_j').$$

For this model we have the following result:

Proposition 1: For model (3.1) consider the learning algorithm (3.4) - (3.5) and an REE $\theta^* = (a^*, b^*, c^*)$ in which all roots of b^* lie inside the unit circle. Then assumptions A and B hold on an open set around θ^*. If all the eigenvalues of the derivative matrix $D\mathcal{T}(\theta^*)$ have real parts less than 1, the learning algorithm converges locally to θ^* in the sense of Theorem 1.

The condition on the eigenvalues of $D\mathcal{T}(\theta^*)$ can be interpreted as an Expectational Stability condition which has been shown to govern the convergence of some adaptive learning algorithms in a variety of models (see, for example, our survey Evans and Honkapohja (1995b)). $T(\theta)$ can be viewed as a mapping from the "perceived law of motion", parameterized by θ, to the implied "actual law of motion" $T(\theta)$ that would be induced by those (fixed) perceptions. Expectational stability is then simply defined as the local stability of θ^* under the differential equation $d\theta/d\tau = T(\theta) - \theta$, where τ is notional time.

The linear dynamic expectations framework (3.1) covers many macroeconomic models. See McCallum (1983) and its references for a number of specific economic models which fit this framework. This section has demonstrated the applicability of Theorem 1 to this class of models, showing how to obtain the local stability conditions for specified solutions and make positive probability statements, without recourse to a projection facility, about convergence from nearby starting points.

We do not formally present instability results in this paper, but it is well-known that one can show convergence with probability zero to an equilibrium which strictly fails to meet the stability condition given in our Proposition.[10] Proposition 1, together with this instability result, is thus the basis for a classification of the REE solutions in this class of models as locally stable or unstable under adaptive learning.

----

10 The relevant technique is to apply Theorem 2 of Ljung (1977). For a closely related example see Evans and Honkapohja (1994a).

b. A scalar example

We now specialize (3.1a-b) to a scalar linear model, so that $\alpha, \beta, \delta, \kappa, \varphi \in \mathbb{R}$ with $\beta \neq 0$, $\delta \neq 0$, $|\varphi| < 1$. This arises in various contexts, see for example the linear-quadratic model in Marcet and Sargent (1989, example e). Grandmont and Laroque (1991) and Grandmont (1994) examined the nonstochastic case $y_t = \beta y^e_{t+1} + \delta y_{t-1}$ (we discuss this case in Section 3c).

The "minimal state variable" solutions are of the form:

$$y_t = a^* + b^* y_{t-1} + c^* w_t + d^* v_t, \tag{3.6}$$

where b^* is a (real) root b of the characteristic equation

$$b^2 - \beta^{-1} b + \beta^{-1} \delta = 0,$$

and $c^* = \kappa[1-\beta(b^*+\varphi)]^{-1}$, $a^* = \alpha^*/(1-\beta-\beta b^*)$ and $d^* = (1-\beta b)^{-1}$. Note that

$$b^* = [1 \pm (1-4\beta\delta)^{1/2}]/2\beta.$$

We will use the notation b^*_+ and b^*_- to denote the two solutions in which the positive radical is added or subtracted, respectively. We assume that neither of the roots is equal to β^{-1} or $\beta^{-1}-\varphi$. With arbitrary values for β and δ the number of real solutions of form (3.6) can be 0, 1 or 2. We will here restrict attention to the case in which there are two real roots.

The analysis of the learning dynamics proceeds in the same way as in the multivariate case. The mapping $T(\theta)$ from the perceived to the actual law of motion takes the form

$$T(\theta) = [(\alpha+\beta a)/(1-\beta b),\ \delta/(1-\beta b),\ (\kappa+\beta\varphi c)/(1-\beta b)]$$

for $\theta' = (a,b,c)$, provided $b \neq \beta^{-1}$. The condition for local convergence is that the eigenvalues of $DT(\theta^*)$ have real parts less than 1. This requires that the following local stability conditions (the "expectational stability" or "E-stability" conditions) be met at the REE:

$$\delta\beta(1-\beta b^*)^{-2} < 1,\ \beta(1-\beta b^*)^{-1} < 1 \text{ and } \varphi\beta(1-\beta b^*)^{-1} < 1.$$

Note that if $0 \leq \varphi < 1$ then the third condition is redundant. To simplify the discussion below we will make this additional assumption.

Before further considering the issue of stability under learning we remark that in fully specified economic models it is often possible to rule out explosive solutions, with $|b| > 1$, on the basis of nonnegativity or transversality conditions. We therefore restrict attention to solutions in which $|b^*| < 1$. The case of "saddlepoint stability", in which one root is smaller than 1 in magnitude and the other root is larger than 1 in magnitude, is frequently encountered. In this case there is a unique nonexplosive solution. However, cases in which both solutions b_+^* and b_-^* are in magnitude less than 1 do also arise.[11]

Figure 1 shows the possibilities in terms of the different regions of the (β,δ) parameter space. In region I, defined by $|\beta + \delta| < 1$, we have the case of saddlepoint stability: $|b_+^*| > 1$ and $|b_-^*| < 1$. In this region the REE (3.6) with $b^* = b_-^*$ is E-stable and therefore locally stable under recursive least squares. In regions II, III and IV, both solutions of the form are nonexplosive, i.e. $|b_-^*| < 1$ and $|b_+^*| < 1$. In region II the b_-^* solution is stable under learning (i.e. locally stable under recursive least squares) while b_+^* is unstable. In region III b_+^* is stable under learning while b_-^* is unstable, and in region IV neither solution is stable under learning. Outside these marked regions either the roots are complex or both real roots are explosive.

FIGURE 1 HERE

In models with multiple equilibria there can in general be nonconvergent paths under least-squares learning. Our central result is that one can nonetheless make positive statements about *local* convergence, even in models with random shocks, *without* imposing a projection facility. If agents' initial estimates of a, b and c are sufficiently close to the stable REE values, then with positive probability there will be convergence to this REE. Furthermore, Corollary 1.1 and Proposition 1.1 imply that, in the case of sufficiently slow adaption, and with starting points near an E-stable solution of the form (3.6), there will be convergence to that solution with probability near 1.

----

[11]For example, the saddlepoint stable case arises in the linear quadratic market model of Sargent (1987, Section XIV, 4 and 6). The magnitude of both roots can be less than 1 if externalities or taxes are introduced into that model as in Sargent (1987, Section XIV.8).

c. Relationship to the Deterministic Case

Grandmont and Laroque (1991) and Grandmont (1994) (henceforth "GLG") have recently analysed the scalar nonstochastic case in detail. They present both stability and instability results, but as is apparent from the title of their joint paper, their emphasis is on the possible instability of learning dynamics.

Some of their results have analogues in the stochastic case. In particular Proposition 2, part 2, of Grandmont and Laroque (1991) provides assumptions yielding local stability for least squares (with finite memory) in the univariate nonstochastic case.[12] Our Proposition 1 generalizes this to multivariate stochastic systems. A major contribution of our paper is precisely to show how it is possible, without the use of an ad hoc projection facility, to state positive convergence results in stochastic models with multiple equilibria.

From our perspective, the instability results of Grandmont and Laroque (1991, Proposition 2, part 1) and Grandmont (1994, Proposition 4.4) reflect the fact that in this model an REE can be only *locally* stable: Convergence of ϕ_t cannot be expected to occur for initial values ϕ_0 sufficiently far from a stable REE ϕ^*. However, the instability results of GLG, e.g. Grandmont's "general instability result", Proposition 4.2, may appear to be more disquieting than this interpretation indicates. For example, Grandmont (1994, p. 27) writes:

> "...if the 'projection facility'...is relatively large, and if expectations matter significantly, then one should get local instability of the actual dynamics with learning for an open set of small initial perturbations. Owing to the discontinuity of the forecasting rule, however, there may also exist here another open set of small initial perturbations generating local convergence."

Thus the instability results of GLG appear to suggest that *any* REE in their model will at best be semistable, since there will be divergence from some nearby initial conditions. How can one reconcile this with our positive probability results, for example our claim that for model specifications in region I of Figure 1, the REE solution of the form (3.6) with $b^* =$

----

[12]In Grandmont (1994) the relevant Proposition is 4.3, second paragraph.

b^*_- is locally stable under adaptive learning, and that for sufficiently low adaption speeds the probability can be made arbitrarily close to 1?

The key lies in noting that GLG examine a nonstochastic model, and in understanding that some aspects of the analysis of learning in nonstochastic systems do not carry over in a natural way to stochastic frameworks. For this discussion we further specialize to the model (1.1)

$$y_t = \beta y^e_{t+1} + \delta y_{t-1} + v_t$$

in order to facilitate a more direct comparison to GLG.[13]

A key difference between the stochastic and nonstochastic versions of the model lies in the REE solutions themselves. Recall that in the stochastic case the minimal state variable solutions are of the form (1.1) which we repeat here for convenience:

$$y_t = b\, y_{t-1} + d\, v_t. \tag{3.7}$$

Assuming $|b|<1$, in the stochastic case the stationary solutions are AR(1) processes with a positive variance and autocorrelation structure

$$\mathrm{cor}(y_t, y_{t-i}) = b^i. \tag{3.8}$$

In a *local* analysis of learning, the *initial data should reflect these time series properties*. In contrast, in the nonstochastic case the only *stationary* solution is $y_t \equiv 0$ which carries no information about the serial correlation of the data. The procedure by GLG of selecting arbitrary initial conditions around 0, which lies behind some of their instability results, leads to a nonlocal analysis from the stochastic viewpoint.[14]

It is worth elaborating on this point. Consider a set of initial data $y_{-L},\ldots,y_0$ generated *near* an REE of the form (3.7). This can be accomplished in several ways, and we specify two natural procedures. One method to generate data near the REE is to use the REE (3.7) and add a small white noise measurement error to each data point. A second procedure would be to generate the initial data using (3.7) but with a small perturbation in the value

----

13 Thus, in (3.1) we have set $\alpha=\kappa=0$.

14 See Honkapohja (1994) for further discussion of the various senses of local stability and instability being employed.

of b from the REE value.

Suppose agents use least squares on data generated by either of these procedures to obtain initial parameter estimates for their learning algorithm. Then it is straightforward to show that the initial estimate of b will be close to the REE value, provided L is sufficiently large.[15] Thus, local analysis of learning for a given REE corresponds to having initial parameters in a neighborhood of that REE. Furthermore, using the expression for J(0) with $\gamma_t=(L+t)^{-1}$ in Theorem 1, it follows that if L is sufficiently large, the probability of convergence to a stable REE is close to 1.

In contrast, much of the instability analysis of GLG is conducted in an arbitrary neighborhood of $y_t=0$. As is apparent, the least squares estimate of b is undefined at (0,...,0) and it is ill-behaved for many points within a neighborhood of (0,...,0). This is, of course, an observation which is fully recognized by GLG and which they exploit. However, we emphasize that "most" points within an neighborhood of (0,...,0) do not constitute data which are "near" the REE, in the precise sense that for most such points (3.8) does not approximately hold[16]. Even if the variance of v_t is made arbitrarily small, there is an essential difference between the data of the stationary REEs in stochastic and nonstochastic models. Only in stochastic models are the dynamic properties revealed by a stationary solution. These properties are entirely lost by the stationary nonstochastic REE solution.[17]

There is, as GLG point out, a "discontinuity" involved in least squares learning in this model. However, from the point of view of learning theory the important discontinuity is at $\mathrm{Var}(v_t)=0$. In stochastic models with $\mathrm{Var}(v_t)>0$ (or with $\kappa\neq 0$ and $\mathrm{Var}(e_t)>0$), the results of this paper show how to obtain robust positive stability results in a local

----

15 For the second procedure this follows from the consistency of least squares. For the first procedure a small measurement error variance generates a small bias in the estimate.

16 For example, if each point of the initial data was given by the nonstochastic stationary equilibrium value of 0 plus a small white noise measurement error, then instead we would have $\mathrm{cor}(y_t, y_{t-i})=0$.

17 In the nonstochastic model these dynamic properties would be revealed only on nonstationary perfect foresight paths.

analysis of least squares learning with no requirement of a projection facility.[18]

4. Global Convergence

It is also possible to use results on probability bounds, such as Theorem 1 above, to obtain global convergence results in models with a unique equilibrium. Theorem 2 below can be used to extend some of the earlier models in the economics literature. For this result it is necessary to introduce stronger assumptions than those of Section 2. First, we strengthen (A.2) and (A.3) to:

(D.1) The functions $H(\theta,x)$ and $\rho_t(\theta,x)$ satisfy for all $\theta,\theta'\in\mathbb{R}^d$, and all $x_1,x_2,x\in\mathbb{R}^k$:

(i) $|H(\theta,x_1)-H(\theta,x_2)| \le L_1(1+|\theta|)|x_1-x_2|(1+|x_1|^{p_1}+|x_2|^{p_1})$,

(ii) $|H(\theta,0)-H(\theta',0)| \le L_2|\theta-\theta'|$,

(iii) $|\partial H(\theta,x)/\partial x-\partial H(\theta',x)/\partial x| \le L_2|\theta-\theta'|(1+|x|^{p_2})$,

(iv) $|\rho_t(\theta,x)| \le C_2(1+|\theta|)(1+|x|^q)$,

for some constants L_1, L_2, p_1, p_2 and q.

The analysis in this section is carried out under the hypothesis that the dynamics for the state variable do not depend on the vector of parameters:

(D.2) The dynamics for X_t satisfy (B.1)-(B.3) and the matrices A and B are independent of θ.

With these assumptions we have a result establishing global convergence in algorithms

----

[18]We remark on one other sense in which the analysis of learning in stochastic and nonstochastic systems is essentially different. The finite memory learning rules studied by GLG cannot converge to REE in the presence of random shocks, since the parameter estimates will be asymptotically noisy. Only decreasing gain learning rules (with $\gamma_t \to 0$) have a chance of convergence to REE in a stochastic system.

with a unique equilibrium:[19]

<u>Theorem</u> <u>2</u>: Under (A.1), (D.1), and (D.2) assume that there exists a unique equilibrium point $\theta^*\in\mathbb{R}^d$ of the associated differential equation $d\theta/d\tau=h(\theta(\tau))$. Suppose that there exists a positive C^2 function $U(\theta)$ on $\mathbb{R}^d$ with bounded second derivatives satisfying

(i) $U'(\theta)h(\theta) < 0$ for $\forall\theta\neq\theta^*$,

(ii) $U(\theta) = 0$ iff $\theta=\theta^*$,

(iii) $U(\theta)\ge\alpha|\theta|^2$ $\forall\theta$ with $|\theta|\ge\rho_0$ for some α, $\rho_0>0$.

Then the sequence θ_n converges $P_{0,x,a}$ a.s. to θ^*.

Although the assumptions of Theorem 2 are quite strong they are sufficient to provide a generalization of the global convergence result of Bray and Savin (1986) to multivariate linear models (see Section 5).

Note that Theorem 2 provides a way for establishing global convergence based on the Lyapunov function of the associated differential equation. It is unnecessary to add on an arbitrarily large projection facility as in Marcet and Sargent (1989a) or Evans and Honkapohja (1995b).

5. An Example of Global Convergence

We consider the multivariate model

$$y_t = \mu + \mathcal{A}y_t^e + \mathcal{C}w_t \quad (5.1a)$$

$$w_t = \mathcal{B}w_{t-1} + v_t, \quad (5.1b)$$

discussed in Evans and Honkapohja (1995b). This generalizes the Bray and Savin (1986) setup. Here y is an n x 1 endogenous vector, w is an observed p x 1 vector of exogenous variables and v is a p x 1 vector of white noise shocks. We assume that all eigenvalues of the p x p

----

[19]In appendix 1 we again treat a somewhat more general case, where X_t is a Markov process independent of θ_{t-1}.

matrix $\mathcal{B}$ lie inside the unit circle, so that w_t is a stationary process. We also assume that v_t and hence w_t has finite moments. y_t^e denotes the (in general nonrational) expectation of y_t held by agents at t-1. The parameters μ, $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are conformable and we assume that I-$\mathcal{A}$ is invertible.

Evans and Honkapohja (1995b) established global convergence to the unique REE of (5.1). However, they did so using Ljung's Theorem 4 and the analysis required the assumption of an (arbitrarily large) projection facility. By using the results of the previous section it is possible to completely dispense with a projection facility and thereby generalize the univariate iid results of Bray and Savin (1986).

Let $z_t' = (1, w_t')$ and $\phi' = (a,b)$. It is easily verified that the model (4.5) has the unique RE solution

$$y_t = \bar{\phi}' z_{t-1} + \eta_t, \tag{5.2}$$

where $\bar{\phi}' = [(I-\mathcal{A})^{-1}\mu, (I-\mathcal{A})^{-1}\mathcal{C}\mathcal{B}]$ and $\eta_t = \mathcal{C}v_t$. It is assumed that at any moment of time agents forecast using the assumed law of motion of the form

$$y_t = \phi_{t-1}' z_{t-1} + \eta_t,$$

where $E_{t-1}\eta_t=0$. Then $y_t^e = \phi_{t-1}' z_{t-1}$ and the actual law is

$$y_t = T(\phi_{t-1})' z_{t-1} + \mathcal{C}v_t, \tag{5.3}$$

where

$$T(\phi)' = (\mu + \mathcal{A}a, \mathcal{A}b + \mathcal{C}\mathcal{B}). \tag{5.4}$$

For the algorithm we specify[20]

$$\phi_t = \phi_{t-1} + \gamma_t R_t^{-1} z_{t-1}(y_t - \phi_{t-1}' z_{t-1})', \tag{5.5a}$$

$$R_t = R_{t-1} + \gamma_t (z_{t-1}z_{t-1}' - R_{t-1}). \tag{5.5b}$$

For some t the matrix R_t may not be invertible. Since this happens only a finite number of times with probability one, R_t can be given an arbitrary value in equation (5.5a) on these occasions (for example, the last invertible value).

----
20 Again we make the additional technical assumption $\limsup_{t\to\infty} (1/\gamma_{t+1})-(1/\gamma_t) < \infty$.

Note that here we allow the estimates at t to depend on y_t, since agents form their forecasts at t-1. The result is:

Proposition 2: Assume that the eigenvalues of $\mathcal{A}$ - I have negative real parts. Suppose that the agents use the learning algorithm (5.3) - (5.5). Then ϕ_t converges a.s. to $\bar{\phi}$ for any initial values.

A model covered by this section is the multivariate version of the Muth market model. For example, Guesnerie (1992, section III.B) considers a specification with two interrelated markets.

6. Conclusions

This paper has applied new methods for analyzing the local stability of learning to stochastic dynamic economic models where multiple equilibria may be present. The combination of multiplicity and random shocks makes local analysis particularly challenging because of the possibility of these shocks moving the state of the system by large amounts. We have shown how to derive local stability conditions and obtain local convergence results without the addition of a "projection facility" which *a priori* constrains estimates to remain in a neighborhood of some particular solution. Our results, giving positive or near 1 probability of convergence from nearby starting points for solutions satisfying a stability condition, provide precise content to the notion of local stability under learning. In combination with known results giving probability of 0 convergence for REE which fail the stability condition, these results can be used to classify REE as locally stable or unstable under adaptive learning.

In the case of a unique REE these methods also provide a means of establishing global convergence (without a projection facility) based on the study of the associated differential equation. This contrasts favorably with previous techniques for establishing global stability and permits the generalization of earlier results.

Finally we remark that, in the case of multiple equilibria, non-local results may be obtainable on the basis of Corollary 1.2 which established a neighborhood of positive probability of convergence. The set of initial conditions from which the system can reach this neighborhood in finite time with positive probability could itself be large. For particular models this set could be examined by numerical methods.

Appendix 1: Recursive Algorithms with a Markovian State Vector

Here we generalize the treatment in the text by allowing the state vector in the algorithm to be a Markov process with possible dependence on the vector of parameters. Such a generalization is valuable, since some economic models require a setup that goes beyond the conditionally linear dynamics used in the text (see Evans and Honkapohja (1996), Kuan and White (1994), and Lettau and Uhlig (1993) for uses of nonlinear state variable dynamics). Moreover, the linear case is most naturally proved using the more general Markovian framework, see Proposition A.1 below. The result is due to Benveniste, Métivier and Priouret (1990)[21], but we follow the companion paper Evans and Honkapohja (1996) and provide assumptions directly on the algorithm and the Markov state process.

Let the algorithm be

$$\theta_t = \theta_{t-1} + \gamma_t H(\theta_{t-1}, X_t) + \gamma_t^2 \rho_t(\theta_{t-1}, X_t),$$

where $\theta_t \in \mathbb{R}^d$ is the vector of parameter estimates and $X_t \in \mathbb{R}^k$ is the observable state vector following a Markov process with a transition probability law $\Pi_\theta(x,A)$ which may depend on θ_{t-1}, i.e. $\mathrm{Prob}[X_t \in A | \theta_{t-1}, X_{t-1}] = \Pi_{\theta_{t-1}}(X_{t-1}, A)$ for Borel sets $A \subset \mathbb{R}^k$. Thus $(X_t, \theta_t)_{t \geq 0}$ is a Markov process.

As in the text we fix an open subset $D \subseteq \mathbb{R}^d$ and maintain assumptions (A.1) - (A.3).

Before stating the alternative assumptions concerning state dynamics, we introduce the following notation and definitions. For any function $f(\theta,x)$ denote by f_θ the mapping $x \to f(\theta,x)$. If $f(\theta,x)$ is differentiable in x we denote by $f'(\theta,x)$ its derivative with respect to x. Also, for any function $g: \mathbb{R}^k \to \mathbb{R}^k$ we define

$$[g]_p = \sup_{x_1 \neq x_2} \frac{|g(x_1) - g(x_2)|}{|x_1 - x_2|(1 + |x_1|^p + |x_2|^p)},$$

and introduce the function space $Li(p) = \{g | \, [g]_p < \infty\}$.

The assumptions on the Markov process for the state variable X_t are:

----
[21] We hereafter refer to this book as BMP.

(C.1) For any compact $Q \subset D$ and any $q>0$ there exists $\mu_q(Q) < \infty$ such that $\forall n$, $x \in \mathbb{R}^k$, $a \in \mathbb{R}^d$:

$$E_{x,a}\{I(\theta_k \in Q, k \leq n)(1 + |X_{n+1}|^q)\} \leq \mu_q(Q)(1 + |x|^q).$$

Here $I(A)$ and $E_{x,a}(.)$ denote, respectively, the indicator function of any set A and the conditional expected value given initial conditions $X_0=x$ and $\theta_0=a$. (C.1) states in essence that the conditional moments of X_n, provided θ_k remains in Q, are uniformly bounded in θ and the bounds are polynomial in the initial state x.[22]

In addition, let $\Pi^n_\theta(x,A) = \text{Prob}_\theta[X_{t+n} \in A | X_t=x]$ denote the transition probability measure n steps ahead for a Markov process X_t with fixed 1-step ahead transition probability $\Pi_\theta(x,A)$. Now assume for any compact $Q \subset D$:

(C.2) $\forall n,\theta$, and $m \geq 0$ $\exists K$: $\int (1+|y|^m)\Pi^n_\theta(x,dy) \leq K(1+|x|^m)$.

(C.3) For some $p>0$ there exist K_1, K_2, q_1 and $\rho_1 < 1$ such that for all functions $g \in Li(p)$, $n \geq 0$, $x_1, x_2 \in R^k$, $\theta, \theta' \in Q$:

$$\text{(i) } \left|\int g(y)\Pi^n_\theta(x_1,dy) - \int g(y)\Pi^n_\theta(x_2,dy)\right| \leq K_1\rho_1^n[g]_p|x_1-x_2|(1+|x_1|^p+|x_2|^p),$$

$$\text{(ii) } \left|\int g(y)\Pi^n_\theta(x,dy) - \int g(y)\Pi^n_{\theta'}(x,dy)\right| \leq K_2[g]_p|\theta-\theta'|(1+|x|^{q_1}).$$

(C.4) For some $p>0$ and for all differentiable functions g with $g' \in Li(p)$ there exists $K_3(g')$ such that $\forall n \geq 0$, $x_1, x_2 \in R^k$, $\theta, \theta' \in Q$:

$$\left|\int g(y)\Pi^n_\theta(x_1,dy) - \int g(y)\Pi^n_\theta(x_2,dy) - \int g(y)\Pi^n_{\theta'}(x_1,dy) + \int g(y)\Pi^n_{\theta'}(x_2,dy)\right| \leq K_3(g')\rho_2^n|\theta-\theta'|(1+|x_1|^{q_2}+|x_2|^{q_2})$$

for constants $\rho_2 < 1$ and q_2 independent of g.

----

[22] This assumption on the moments and assumption (A.1) on the sequence of gains can be relaxed at the cost of substantial additional technical detail, see BMP, Chapter 3 of Part II.

With the assumptions A and C replacing A and B Theorem 1 holds as before. See the companion paper, Evans and Honkapohja (1995c), for a complete exposition and proof. Proposition A.1 below shows that assumptions (B.1)-(B.3) in the text are a special case of (C.1)-(C.4).

Finally, we remark that Theorem 2 on global convergence also holds in the case of Markovian state dynamics X_t verifying:

(D.2') The transition probability law $\Pi(x,dy)$ of the Markov process for X_t is independent of θ and satisfies:

(i) $\forall n, m \geq 0$ $\exists K$: $\int (1+|y|^m)\Pi^n(x,dy) \leq K(1+|x|^m)$,

(ii) For $\forall p \geq 0$ $\exists K_1$ and $\rho < 1$ such that for all functions $g \in Li(p)$, $n \geq 0$, $x_1, x_2 \in R^k$:

$$\left|\int g(y)\Pi^n(x_1,dy) - \int g(y)\Pi^n(x_2,dy)\right| \leq K\rho^n[g]_p|x_1-x_2|(1+|x_1|^p+|x_2|^p).$$

Appendix 2: Proofs of results.

<u>Proof of Lemma</u>: By the martingale convergence theorem assumptions (B.2) and (B.3) imply the existence of the limiting random variable $V_\infty(\theta) = \lim_{n\to\infty} V_n$ (almost surely and in L_p) for any given θ, where

$$V_n = \Sigma_{k=1}^n A^k(\theta)B(\theta)W_k.$$

(Clearly, V_n is bounded given the assumptions B.) By symmetry it is seen that for all n V_n has the same distribution as n-th stage

$$U_n = \Sigma_{k=1}^n A^{n-k}(\theta)B(\theta)W_k, \quad \text{(a1)}$$

of the process in (B.1). The probability measure of V_∞ is the invariant measure of $\tilde{X}_t(\theta)$. By symmetry we have for any continuous function g satisfying $|g(x)| \leq C(1+|x|^q)$ and any initial x

$$E(g(A^n(\theta)x + U_n) = E(g(A^n(\theta)x + V_n). \quad \text{(a2)}$$

$A^n(\theta)x \to 0$ implies $\lim_{n\to\infty} E(g(A^n(\theta)x + U_n) = E(g(V_\infty))$. Using (A.2i) we have that $h(\theta)$ is well-defined.

The Lipschitz property is proved as follows. First note that (A.3) clearly implies that

$$|H(\theta,x)-H(\theta',x)| \le L_2|\theta-\theta'|(1+|x|) \text{ and}$$

$$|H(\theta,x)-H(\theta',x')| \le |H(\theta,x)-H(\theta',x)| + |H(\theta,x')-H(\theta',x')|$$

$$\le C_1|\theta-\theta'|(1+|x|)+C_2|x-x'|.$$

Therefore,

$$|EH(\theta,\bar{X}_t(\theta))-EH(\theta',\bar{X}_t(\theta'))| \le$$

$$E|H(\theta,\bar{X}_t(\theta))-H(\theta',\bar{X}_t(\theta'))| \le$$

$$E[C_1|\theta-\theta'|(1+|\bar{X}_t(\theta)|)+C_2|\bar{X}_t(\theta)-\bar{X}_t(\theta')|] \le$$

$$C_1|\theta-\theta'|E(1+|\bar{X}_t(\theta)|)+C_2E|\bar{X}_t(\theta)-\bar{X}_t(\theta')|.$$

In the last expression lim $E(1+|\bar{X}_t(\theta)|)$ is bounded by a constant. Since $A(\theta)$ and $B(\theta)$ are assumed Lipschitz (see (B.2)) we also have that

$$\lim E\ |\bar{X}_t(\theta)-\bar{X}_t(\theta')| \le K|\theta-\theta'|$$

for some constant K. Hence $h(\theta)$ is Lipschitz.

To establish Theorem 1 of the text we show that Assumptions B of the text imply the assumptions C for the case of general Markovian state dynamics, Appendix 1.

Proposition A.1: If (B.1) - (B.3) hold, then X_t satisfies (C.1) - (C.4) for any $p>0$.

Proof of Proposition A.1: We develop here the main steps, but refer to BMP for some lengthy details. First, it may be shown that process $U_n(\theta)$, defined in (a1), satisfies $\|U_n(\theta)\|_p \le K$ and $\|U_n(\theta) - U_n(\theta')\|_p \le K^*|\theta - \theta'|$ for all p and some constants K, K* (see BMP pp.266-267).

Next we note that (B.3) implies $|A^n(\theta)| \le \rho^n$. (C.1) is then immediate from (B.2) and (B.3). (C.2) follows at once from applying equation (a2) to the function $g(y)=1+|y|^m$ and

using the bounds in (B.2) and (B.3).

We easily obtain (C.3)(i) from the definition of $[g]_p$ using the inequality

$$|E[g(A^n(\theta)x_1 + U_n(\theta)) - g((A^n(\theta)x_2 + U_n(\theta))]| \le$$

$$[g]_p|A^n(\theta)||x_1-x_2|E[1 + |A^n(\theta)x_1 + U_n(\theta)|^p + |A^n(\theta)x_2 + U_n(\theta)|^p].$$

To provc (C.3)(ii) we note that

$$|E[g(A^n(\theta)x + U_n(\theta)) - g((A^n(\theta')x + U_n(\theta'))]| \le$$

$$[g]_pE\{[|A^n(\theta)-A^n(\theta')||x|+|U_n(\theta)-U_n(\theta')|]\ [1+|A^n(\theta)x+U_n(\theta)|^p+|A^n(\theta')x+U_n(\theta')|^p]\}$$

to which we apply Cauchy-Schwarz inequality. Using (B.2) and (B.3) the resulting second term is polynomially bounded in $|x|$. The resulting first term is is simply $\||A^n(\theta)-A^n(\theta')||x| + |U_n(\theta) - U_n(\theta')|\|_2$. In the beginning of the proof it was noted that U_n satisfies a Lipschitz condition. Since $A(\theta)$ is assumed also to satisfy a Lipschitz condition and $|A^n(\theta)| \le \rho^n$ the whole expression is bounded by an expression of the form $C|\theta-\theta'|(1+|x|^q)$ which proves C(3)(ii).

We omit the proof of (C.4): for the lengthy details see BMP, p.269 (we remark that $K_3(g')=N_p(g')$ in BMP's notation).

Remark on the Proof of Theorem 1: In view of Proposition A.1 it suffices to prove that Theorem 1 holds with assumptions A of section 2 and C of Appendix 1. We prove this result in the companion paper, Evans and Honkapohja (1996).

Proof of Corollary 1.1: Immediate from Theorem 1.

Proof of Corollary 1.2: Fix compact $Q \subset D$. Using Theorem 1 there is $\bar{n}$ such that $P_{\bar{n},x,a}\{\theta_t \to \theta^*\} \ge q > 0$. Substituting recursively in the algorithm it follows that $\theta_{\bar{n}} = Z(\theta_0, W_0,\ldots,W_{\bar{n}-1},X_0)$ is a continuous function, because H and ρ_k are continuous. Furthermore, $Z(\theta^*,w^*,\ldots,w^*,x^*)=0$. There exists $\zeta>0$ such that $\theta_{\bar{n}} \in D$ for $|\theta_0-\theta^*|<\zeta$, $|W_i-w^*|<\zeta$, $i=1,\ldots,\bar{n}$, and $|X_0-x^*|<\zeta$. Because of the assumption of positive density, this event has a positive probability. The conclusion follows.

Proof of Proposition 1: In order to write the algorithm (3.5) in the form (2.1) we set $S_{t-1} \equiv R_t$ and define $X_t' = (1, y_{t-1}', w_t', y_{t-2}', w_{t-1}', v_{t-1}')$ and θ_t as a vector composed of the elements of ϕ_t and S_t. Note that $z_t' = (1, y_{t-1}, w_t)$ is a function of X_t. Equation (3.5a) forms directly a component of function $H(\theta, X)$ in (2.1), while (3.5b) can be written in the form

$$S_t = S_{t-1} + \gamma_t(z_t z_t' - S_{t-1}) + \frac{\gamma_{t+1} - \gamma_t}{\gamma_t^2}\gamma_t^2 (z_t z_t' - S_{t-1}), \tag{a3}$$

which defines implicitly the remaining part of $H(\theta,X)$ of form (2.1) with $\rho_t(S_{t-1}, z_t) = \frac{\gamma_{t+1}-\gamma_t}{\gamma_t^2}(z_t z_t' - S_{t-1})$.

In order to define domain D of the algorithm let ϕ^* be a fixed point of $T(\phi)$, and assume that the eigenvalues of b^* are strictly inside the unit circle, where $\phi^{*\prime} = (a^*, b^*, c^*)$. Define $z_t(\phi)' = (1, y_{t-1}'(\phi), w_t')$, where we have $y_{t-1}(\phi) = T(\phi)' z_{t-1}(\phi)$. Then $z_t(\phi)$ is a stationary process for all ϕ sufficiently near ϕ^*. Let $M_z(\phi) = E(z_t(\phi) z_t(\phi)')$, and note that $S^* = E[z_t(\phi^*) z_t(\phi^*)']$ is positive definite. Next choose an open set $\hat{D}$ around (ϕ^*, S^*) so that for all $(\phi, S) \in \hat{D}$:

(i) ϕ^* is the unique fixed point of T,

(ii) for some $\varepsilon > 0$ $\det(S) \geq \varepsilon > 0$,

(iii) $(I - \beta b)$ is invertible, and

(iv) the roots of b are bounded strictly inside the unit circle.

With this construction it is straightforward to show that conditions (A.1) - (A.3) hold.[23] To verify conditions (B.1)-(B.3) we note that the system for the state vector X_t is of form (B.1) with

----

23 For verification of (A.2ii) note that $(\gamma_{t+1} - \gamma_t)/\gamma_t^2 \leq 1/\gamma_{t+1} - 1/\gamma_t$ which is bounded.

$$W_{t-1} = \begin{pmatrix} 1 \\ v_{t-1} \\ e_t \end{pmatrix}, \quad A(\theta_{t-1}) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ T_a(\phi_{t-1}) & T_b(\phi_{t-1}) & T_c(\phi_{t-1}) & 0 & 0 & 0 \\ 0 & 0 & \varphi & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\text{and } B(\theta_{t-1}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & W(b_{t-1}) & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

where $T(\phi)' = (T_a(\phi), T_b(\phi), T_c(\phi))$. Condition (B.2) then holds by the assumption of iid shocks with bounded moments. (B.3) is also satisfied, provided the neighborhood $\hat{D}$ around the point (ϕ^*, S^*) is made sufficiently small, since the nonzero eigenvalues of $A(\theta)$ are the eigenvalues of φ and $T_b(\phi)$.

In the associated differential equation $d\theta/d\tau = h(\theta)$ the right-hand side can be written the form

$$h_\phi(\phi, S)' = S^{-1} M_z(\phi)(T(\phi) - \phi)$$

$$h_S(\phi, S) = M_z(\phi) - S.$$

If $D\mathcal{T}(\phi^*)$ has all eigenvalues with real parts less than 1, then it can be shown (Marcet and Sargent 1989a, Proposition 3) that the associated differential equation is locally asymptotically stable at (ϕ^*, S^*). Using Proposition 1 in section 2 one may choose $D = \mathrm{int}(K(c))$ so small that $D \subset \hat{D}$, and the conclusion follows.

Proof of Theorem 2: (Remarks) The essence of the argument is to show that the sequence θ_n is bounded a.s. We omit the proof of Theorem 2 which follows step by step the proof of Theorem 17 of BMP, p.239 with the following modifications. First, (D.1) and (D.2) imply that the conclusions of Theorem 7 of BMP, p.265 are satisfied which in turn yields conditions (1.9.1)-(1.9.6) of BMP, p.239 with $\lambda = 1$. Second, when BMP apply their Proposition 7 one can use our Theorem 1.

Proof of Proposition 2: Mimicing the proof of Proposition 1, our algorithm is given by (5.3), (5.4), (5.5a) with R_t replaced by S_{t-1}, and writing (5.5b) in the form (a3) with the new definitions of θ, z and y given above. For global analysis it is necessary to introduce a modified algorithm, since the covariance matrix S_t could fail to be invertible. This modified algorithm will coincide with the original one after a finite (random) time.

Consider first (a3). It satisfies the conditions for Theorem 2 with the associated differential equation

$$h_2(S) = M_z - S, \tag{a4}$$

where $M_z = \lim_{t\to\infty} E(z_t z_t')$ which under our assumptions is positive definite. The corresponding Lyapunov function is $U(S)=\|S-M_z\|^2$. Thus S_t converges a.s. to M_z from any starting point.

Introduce a neighborhood $\mathcal{N}$ of M_z such that S^{-1} exists whenever $S\in\mathcal{N}$. It is possible to construct a bounded regular function u(S) from the space of (p+1)x(p+1) matrices to the subspace of positive definite matrices and such that $u(S)=S^{-1}$ on $\mathcal{N}$. The modified algorithm is obtained by replacing (5.5a) with

$$\theta_t = \theta_{t-1} + \gamma_t u(S_{t-1})z_{t-1}(T(\theta_{t-1})'z_{t-1} + \mathcal{E}v_t - \theta_{t-1}'z_{t-1})'.$$

The assumptions (D.1) and (D.2) are easily verified. The associated differential equation takes the form

$$d\theta/d\tau=(h_1(\theta,S),\ h_2(S)),$$

where $h_1(\theta,S) = u(S)M_z[(\mathcal{A}-I)(\theta'-\bar{\theta}')]'$ and $h_2(S)$ given by (a4). This differential equation is globally asymptotically stable, since the eigenvalues of $\mathcal{A}$-I have negative real parts. Moreover, stability is exponential. It follows (see Hahn 1967) that there exists a C^2 Lyapunov function $W(\theta)$ satisfies $W(\theta) \geq \alpha|\theta|^2$. Finally, to ensure a Lyapunov function $U(\theta)$ with bounded second derivatives we set $U(\theta) = \psi(W(\theta))$, where the transformation ψ satisfies $\psi(0)=0$, $\psi'(t)>0$, $\lim_{t\to\infty}\psi(t)=+\infty$, and is such that the derivatives of ψ tend to zero sufficiently rapidly at infinity. With $U(\theta)$ the remaining requirements of Theorem 2 are

clearly satisfied.

REFERENCES:

Benassy, J.P. and M. Blad (1989), "On Learning and Rational Expectations in an Overlapping Generations Model", Journal of Economic Dynamics and Control, 13, 379-400.

Benveniste, A., M. Métivier and P. Priouret (1990), Adaptive Algorithms and Stochastic Approximations, New York: Springer-Verlag.

Bray, M. and N.E. Savin (1986), "Rational Expectations Equilibria, Learning, and Model Specification", Econometrica 54, 1129-1160.

Evans, G.W. and S. Honkapohja (1994a), "Learning, Convergence and Stability with Multiple Rational Expectations Equilibria", European Economic Review, 38, 1071-1098.

Evans, G.W. and S. Honkapohja (1994b), "On the Local Stability of Sunspot Equilibria under Adaptive Learning Rules", Journal of Economic Theory, 64, 142-161.

Evans, G.W. and S. Honkapohja (1995a), "Local Convergence of Recursive Learning to Steady States and Cycles in Stochastic Nonlinear Models", Econometrica, 63, 195-206.

Evans, G.W. and S. Honkapohja (1995b), "Adaptive Learning and Expectational Stability: An Introduction", in Kirman, A, and M. Salmon, eds., Learning and Rationality in Economics, Oxford: Basil Blackwell, 102-126.

Evans, G. W. and S. Honkapohja (1996), "Convergence of Learning Algorithms without a Projection Facility", mimeo.

Fourgeaud C., C. Gourieroux and J. Pradel (1986), "Learning Procedures and Convergence to Rationality", Econometrica, 54, 845-868.

Gantmacher, F. R. (1959), The Theory of Matrices, vol. 1, New York: Chelsea.

Grandmont, J.-M. (1994), "Expectations Formation and Stability of Large Socioeconomic Systems", CEPREMAP Discussion Paper No. 9424.

Grandmont, J.-M. and G. Laroque (1991), "Economic Dynamics with Learning: Some Instability Examples", in Equilibrium Theory and Applications, Proceedings of the Sixth International Symposium in Economic Theory and Econometrics, ed. by W. A. Barnett et. al. Cambridge: Cambridge University Press, pp.247-273.

Guesnerie, R. (1992), "An Exploration of the Eductive Justifications of the Rational Expectations Hypothesis", American Economic Review, 82, 1254-1278.

Hahn, W. (1967), Stability of Motion, Berlin: Springer Verlag.

Honkapohja, S. (1994), "Expectations Driven Nonlinear Business Cycles: Comments", in Measuring and Interpreting Business Cycles, ed. by V. Bergström and A. Vredin, Oxford: Clarendon Press, 256-262.

Horn, R. A. and C. R. Johnson (1991), Topics in Matrix Analysis, Cambridge, U.K.: Cambridge University Press.

Kuan, C.-M. and H. White (1994), "Adaptive Learning with Nonlinear Dynamics Driven by Dependent Processes", Econometrica, 62, 1087-1114.

Lettau, M. and H. Uhlig (1993), "Rules of Thumb and Dynamic Programming", mimeo, CentER, Tilburg University.

Ljung, L. (1977), "Analysis of Recursive Stochastic Algorithms", IEEE Transactions on Automatic Control AC-22, 551-575.

Marcet, A. and T.J. Sargent (1989a), "Convergence of Least Squares Learning Mechanisms in Self-Referential Stochastic Models", Journal of Economic Theory 48, 337-368.

Marcet A. and T.J. Sargent (1989b), "Convergence of Least Squares Learning in Environments with Hidden State Variables and Private Information", Journal of Political Economy, 97, 1306-1322.

McCallum B.T. (1983), "On Nonuniqueness in Rational Expectations Models: An Attempt at Perspective", Journal of Monetary Economics, 11, 134-168.

Moreno, D. and M. Walker (1994), "Two Problems in Applying Ljung's "Projection Algorithms" to the Analysis of Decentralized Learning", Journal of Economic Theory, 62, 420-427.

Sargent, T.J. (1987), Macroeconomic Theory, 2nd ed., Boston: Academic Press.

Sargent, T.J. (1993), Bounded Rationality in Macroeconomics, Oxford: OUP.

Woodford, M. (1990), "Learning to Believe in Sunspots", Econometrica 58, 277-307.

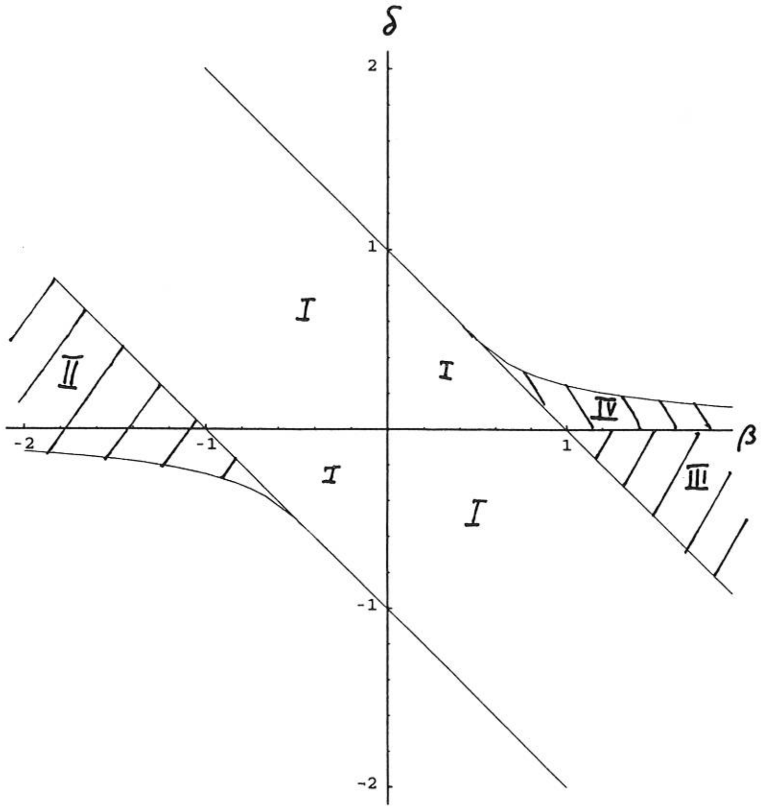

FIGURE I

CES Working Paper Series

---

55 Robert S. Chirinko, Non-Convexities, Labor Hoarding, Technology Shocks, and Procyclical Productivity: A Structural Econometric Approach, January 1994

56 A. Lans Bovenberg and Frederick van der Ploeg, Consequences of Environmental Tax Reform for Involuntary Unemployment and Welfare, February 1994

57 Jeremy Edwards and Michael Keen, Tax Competition and Leviathan, March 1994

58 Clive Bell and Gerhard Clemenz, The Desire for Land: Strategic Lending with Adverse Selection, April 1994

59 Ronald W. Jones and Michihiro Ohyama, Technology Choice, Overtaking and Comparative Advantage, May 1994

60 Eric L. Jones, Culture and its Relationship to Economic Change, May 1994

61 John M. Hartwick, Sustainability and Constant Consumption Paths in Open Economies with Exhaustible Resources, June 1994

62 Jürg Niehans, Adam Smith and the Welfare Cost of Optimism, June 1994

63 Tönu Puu, The Chaotic Monopolist, August 1994

64 Tönu Puu, The Chaotic Duopolists, August 1994

65 Hans-Werner Sinn, A Theory of the Welfare State, August 1994

66 Martin Beckmann, Optimal Gambling Strategies, September 1994

67 Hans-Werner Sinn, Schlingerkurs - Lohnpolitik und Investitionsförderung in den neuen Bundesländern, September 1994

68 Karlhans Sauernheimer and Jerome L. Stein, The Real Exchange Rates of Germany, September 1994

69 Giancarlo Gandolfo, Pier Carlo Padoan, Giuseppe De Arcangelis and Clifford R. Wymer, The Italian Continuous Time Model: Results of the Nonlinear Estimation, October 1994

70 Tommy Staahl Gabrielsen and Lars Sørgard, Vertical Restraints and Interbrand Competition, October 1994

71 Julia Darby and Jim Malley, Fiscal Policy and Consumption: New Evidence from the United States, October 1994

72 Maria E. Maher, Transaction Cost Economics and Contractual Relations, November 1994

73 Margaret E. Slade and Henry Thille, Hotelling Confronts CAPM: A Test of the Theory of Exhaustible Resources, November 1994

74 Lawrence H. Goulder, Environmental Taxation and the "Double Dividend": A Reader's Guide, November 1994

75 Geir B. Asheim, The Weitzman Foundation of NNP with Non-constant Interest Rates, December 1994

76 Roger Guesnerie, The Genealogy of Modern Theoretical Public Economics: From First Best to Second Best, December 1994

77 Trond E. Olsen and Gaute Torsvik, Limited Intertemporal Commitment and Job Design, December 1994

78 Santanu Roy, Theory of Dynamic Portfolio Choice for Survival under Uncertainty, July 1995

79 Richard J. Arnott and Ralph M. Braid, A Filtering Model with Steady-State Housing, April 1995

80 Vesa Kanniainen, Price Uncertainty and Investment Behavior of Corporate Management under Risk Aversion and Preference for Prudence, April 1995

81 George Bittlingmayer, Industry Investment and Regulation, April 1995

82 Richard A. Musgrave, Public Finance and Finanzwissenschaft Traditions Compared, April 1995

83 Christine Sauer and Joachim Scheide, Money, Interest Rate Spreads, and Economic Activity, May 1995

84 Jay Pil Choi, Preemptive R&D, Rent Dissipation and the "Leverage Theory", May 1995

85 Stergios Skaperdas and Constantinos Syropoulos, Competing for Claims to Property, July 1995

86 Charles Blackorby, Walter Bossert and David Donaldson, Intertemporal Population Ethics: Critical-Level Utilitarian Principles, July 1995

87 George Bittlingmayer, Output, Political Uncertainty, and Stock Market Fluctuations: Germany, 1890-1940, September 1995

88 Michaela Erbenová and Steinar Vagstad, Information Rent and the Holdup Problem: Is Private Information Prior to Investment Valuable?, September 1995

89 Dan Kovenock and Gordon M. Phillips, Capital Structure and Product Market Behavior: An Examination of Plant Exit and Investment Decisions, October 1995

90 Michael R. Baye, Dan Kovenock and Casper de Vries, The All-pay Auction with Complete Information, October 1995

91 Erkki Koskela and Pasi Holm, Tax Progression, Structure of Labour Taxation and Employment, November 1995

92 Erkki Koskela and Rune Stenbacka, Does Competition Make Loan Markets More Fragile?, November 1995

93 Koji Okuguchi, Effects of Tariff on International Mixed Duopoly with Several Markets, November 1995

94 Rolf Färe, Shawna Grosskopf and Pontus Roos, The Malmquist Total Factor Productivity Index: Some Remarks, November 1995

95 Guttorm Schjelderup and Lars Sørgard, The Multinational Firm, Transfer Pricing and the Nature of Competition, November 1995

96 Guttorm Schjelderup, Kåre P. Hagen and Petter Osmundsen, Internationally Mobile Firms and Tax Policy, November 1995

97 Makoto Tawada and Shigemi Yabuuchi, Trade and Gains from Trade between Profit-Maximizing and Labour-Managed Countries with Imperfect Competition, December 1995

98 Makoto Tawada and Koji Shimomura, On the Heckscher-Ohlin Analysis and the Gains from Trade with Profit-Maximizing and Labour-Managed Firms, December 1995

99 Bruno S. Frey, Institutional Economics: What Future Course?, December 1995

100 Jean H. P. Paelinck, Four Studies in Theoretical Spatial Economics, December 1995

101 Gerhard O. Orosel and Ronnie Schöb, Internalizing Externalities in Second-Best Tax Systems, December 1995

102 Hans-Werner Sinn, Social Insurance, Incentives and Risk Taking, January 1996

103 Hans-Werner Sinn, The Subsidiarity Principle and Market Failure in Systems Competition, January 1996

104 Uri Ben-Zion, Shmuel Hauser and Offer Lieberman, A Characterization of the Price Behaviour of International Dual Stocks: An Error Correction Approach, March 1996

105 Louis N. Christofides, Thanasis Stengos and Robert Swidinsky, On the Calculation of Marginal Effects in the Bivariate Probit Model, March 1996

106 Erkki Koskela and Ronnie Schöb, Alleviating Unemployment: The Case for Green Tax Reforms, April 1996

107 Vidar Christiansen, Green Taxes: A Note on the Double Dividend and the Optimum Tax Rate, May 1996

108 David G. Blanchflower and Richard B. Freeman, Growing Into Work, May 1996

109 Seppo Honkapohja and George W. Evans, Economic Dynamics with Learning: New Stability Results, May 1996