Macroeconomic Dynamics, **7**, 2003, 119–139. Printed in the United States of America.
DOI: 10.1017.S1365100502010325

STATISTICAL LEARNING WITH TIME-VARYING PARAMETERS

Bruce McGough
Oregon State University

In their landmark paper, Bray and Savin note that the constant-parameters model used by their agents to form expectations is misspecified and that, using standard econometric techniques, agents may be able to determine the time-varying nature of the model's parameters. Here, we consider the same type of model as employed by Bray and Savin except that our agents form expectations using a perceived model with parameters that vary with time. We assume agents use the Kalman filter to form estimates of these time-varying parameters. We find that, under certain restrictions on the structure of the stochastic process and on the value of the stability parameter, the model will converge to its rational expectations equilibrium. Further, the restrictions on the stability parameter required for convergence are identical to those found by Bray and Savin.

Keywords: Bounded Rationality, Model Misspecification, Time-Varying Parameters, Kalman Filter

1. INTRODUCTION

Modern stochastic macroeconomic models typically include, among the factors governing their dynamics, dependence upon the predicted values of endogenous variables. The standard method of analysis of such models comes from the theory of rational expectations. According to this theory, economic agents are assumed to form predictions using conditional mathematical expectations; when these conditional expectations are formed with respect to the distributions of the actual stochastic processes generating the data, the economy is said to be in a rational expectations equilibrium (REE). This notion of equilibrium is well established as the discipline's benchmark; however, it is not without criticism. Evans and Honkapohja (1998, p. 453), note that "rational expectations . . . assumes that agents know the true economic model generating the data and implicitly assumes coordination of expectations by the agents." Further, examples of macroeconomic models with multiple rational expectations equilibria are abundant; the theory gives no indication as to which equilibrium is likely to govern the behavior of the economy. To address these criticisms, some economists choose to weaken the notion of rationality. Instead of assuming agents know the true economic model generating the

I would like to thank George Evans for many valuable discussions. All errors are mine. Address correspondence to: Bruce McGough, Department of Economics, Oregon State University, 303 Ballard Extension Hall, Office 315, Corvallis, OR 97331-3612, USA; e-mail: bruce.mcgough@orst.edu.

 1365-1005/03 $12.00

data, researchers assume agents are boundedly rational. The manifestation of this assumption in models of statistical learning is that agents know the structure of the rational expectations equilibrium and form estimates of the relevant parameters adaptively, using statistical algorithms. With these estimates, agents form their expectations of the values of the endogenous variables. One can then consider whether the economy eventually approaches a rational expectations equilibrium: That is, do the parameter estimates eventually converge (in some probabilistic sense) to the corresponding rational expectations equilibrium parameter values?

The first authors to consider such a model were Bray and Savin. In their landmark paper [Bray and Savin (1986)], they showed that if the economy is governed by a simple cobweb model, and if agents have a perceived model of the same linear functional form as the REE, and if agents estimate the parameters of this model using ordinary least squares, then the estimates indeed converge to the associated REE values, for appropriate values of the stability parameter. Their method of proof was quite technical and based on the theory of Martingales. Marcet and Sargent (1989) were able to extend this result to more general linear models using Ljung's (1977) theory on recursive stochastic algorithms. Evans and Honkapohja (2001) derived a Ljung-type result designed specifically for application to economic models and subsequently extended these convergence results to multivariate linear models. Much further work has been done. For a survey and brief history, see Evans and Honkapohja (1998).

All the results mentioned in the preceding paragraph are derived assuming agents use OLS as their method of estimation. Implicit then is the presumption by agents that the parameters of the model are constant. However, since the estimates themselves necessarily affect the true values of the parameters, these constant parameter beliefs by agents are erroneous. Bray and Savin knew this to be a concern and used simulations to consider whether agents could detect the time-varying nature of the parameters. They found that, for certain initial conditions and parameter values, agents may in fact determine that the model is misspecified. Since it may be possible for agents to realize that the parameters are not constant, it is important to analyze models in which agents believe the parameter values vary with time, hence the topic of this paper.

The behavior of an economic model based on agents with time-varying parameter beliefs has been considered by Bullard (1992). Bullard used a general linear reduced-form model, of which the above-mentioned cobweb model is a special case, to show that if agents believe that the parameters of the perceived model follow a random walk with i.i.d. noise term, then the economy never converges to the REE. This result is not surprising: If agents believe that the parameters of the economy will not settle down, then their estimates of those parameters will not settle down because the agents will always attribute some of the noise in the model to movement in the parameter values. We conclude that, for convergence to a rational expectations equilibrium to occur, the agents must believe that the conditional variance of the time-varying parameters decreases to zero.[1] This is a natural assumption for the agents to make. In particular, if agents initially use

OLS to form their estimates, then the results of Bray and Savin tell us that the conditional variance of the process describing the actual parameters will decrease to zero.[2]

In this paper, we analyze the asymptotic behavior of an economy described by a simple cobweb model with agents who have time-varying parameter beliefs. We find that if they believe the parameters of the economy follow a random walk, and if the conditional variance of this random walk decreases rapidly enough, then convergence to the rational expectations equilibrium obtains for appropriate values of the stability parameter.

This paper is organized as follows: Section 2 begins with a review of the simple cobweb model and results of Bray and Savin and then presents the modification of the model that allows for time-varying parameter beliefs. A change of variables is presented which allows for simpler analysis of the stochastic processes. The main result of the paper ends the section. In Section 3, a more general cobweb model is considered and tools from the theory of stochastic approximation are used to show convergence in this case. A connection with E-stability is also discussed. Section 4 concludes. Most of the technical proofs are relegated to the appendices.

2. BRAY AND SAVIN'S COBWEB MODEL

2.1. Constant-Parameter Beliefs

In this section, we consider the same cobweb model as analyzed by Bray and Savin. The reduced form of this model is

$$y_t = x_t' m + a E_t^* y_t + \nu_t, \tag{1}$$

where y_t is the endogenous variable, $x_t \in R^n$ is an exogenous i.i.d. process observed at time t, the first component of which is 1, ν_t is an unobserved white-noise shock, and $E_t^* y_t$ is the agents' expectation of the value of y_t formed using information up to and including time t. This reduced form may be obtained by modeling a single competitive market with stochastic linear demand and supply which is derived from firms that face quadratic costs and a production lag. A given firm's supply decision is made before stochastic demand is realized, and thus the decision is based on expected equilibrium price. Equating market supply and demand yields equation (1). The parameter a is the ratio of the slopes of supply and demand, and so, provided demand slopes downward and supply slopes upward, we have $a < 0$.[3]

The model is closed by specifying the form of the expectations operator. Provided agents behave rationally,

$$E_t^* y_t = E(y_t \mid \Omega_t),$$

where Ω_t represents the agents' information set. The unique REE, that is, the final form of the model consistent with the assumption of rationality, is then easily computed to be

$$y_t = x_t'\left(\frac{m}{1-a}\right) + \nu_t.$$

To weaken the assumption of rationality and subsequently incorporate learning into their model, Bray and Savin postulate that agents believe that

$$y_t = x_t'\beta + \varepsilon_t$$

but are unaware of the value of β. Further, Bray and Savin assume their agents use OLS to estimate β and then use this estimate to form their expectations. Specifically, let b_t be the OLS estimate of β using data $(x_1, y_1), \ldots, (x_t, y_t)$. Then, $E_t^* y_t = x_t' b_{t-1}$. Agents' expectations feed back into the reduced-form model (1) to yield the actual data-generating process

$$y_t = x_t'(m + ab_{t-1}) + \nu_t. \quad \textbf{(2)}$$

Notice that the parameter modifying x_t, namely $(m + ab_{t-1})$, is time dependent, contrary to the assumption of the agents.

To complete their analysis, Bray and Savin use recursive least squares, together with true process (2), to write the sequence of estimators, b_t, as

$$b_t = \left(I + (a-1)\frac{1}{t}V_t x_t x_t'\right)b_{t-1} + \frac{1}{t}V_t x_t x_t' m + \frac{1}{t}V_t x_t \nu_t, \quad \textbf{(3)}$$

where $V_t = t(\sum_{i=1}^{t} x_i x_i')^{-1}$. Their main result is as follows.

THEOREM 1 [Bray and Savin (1986)]. *If* $a < 1$, *then* $b_t \to m/(1-a)$ *almost surely.*

It is important for our work to observe that the proof of this theorem does not rely on the structure of the process V_t, but only that it converges to $(Ex_t x_t')^{-1}$ almost surely.

2.2. Time-Varying Parameter Beliefs

In this section we alter the model of Bray and Savin by allowing the parameters of the agents' perceived model to vary with time. We then attempt to analyze the resulting asymptotic behavior of the economy. This analysis requires imposing a structure on the believed process describing the time-varying parameters. Here we consider the process to be a random walk with potentially variable conditional variance. The reasons for choosing a random walk are fourfold: first, a random walk is a standard model of time-varying parameters; second, it is consistent with the learning literature [see, e.g., Bullard (1992)]; third, if the conditional variance is zero, then the random walk reduces to the constant-parameter model considered by Bray and Savin; and fourth, its simplicity allows for analytic tractability.

We modify Bray and Savin's model as follows: Assume agents believe

$$y_t = \beta_t' x_t + \epsilon_t$$

$$\beta_{t+1} = \beta_t + \eta_t$$

and that x_t and β_t are independent. We assume, for technical reasons, that $\text{var}(\eta_t) = \hat{\sigma}_t^2 I$. Notice that agents' beliefs form a linear state-space model and thus the Kalman filter is a natural estimator for β_{t+1}. Denote by b_t the estimate of β_{t+1} using data available at time t. The recursions for the filter are given by

$$K_t = P_{t-1}\left[\sigma^2 + x_t' P_{t-1} x_t\right]^{-1}, \quad \textbf{(4)}$$

$$b_t = b_{t-1} + K_t x_t [y_t - b_{t-1}' x_t], \quad \textbf{(5)}$$

$$P_t = P_{t-1} - K_t x_t x_t' P_{t-1} + \text{var}(\eta_t). \quad \textbf{(6)}$$

Here K_t represents the Kalman gain, b_t is the linear projection of β_{t+1} on time-t variables, and P_t is the mean squared error of the estimator b_t conditional on the realizations of x. Notice that if x_t and u_t are normal, b_t is the optimal estimator. For details, see Brockwell and Davis (1987).

Given the agents' beliefs and estimator, we have $E_t y_t = b_{t-1}' x_t$.[4] Inserting this into equation (1) yields the true data-generating process

$$y_t = x_t'(m + a b_{t-1}) + \nu_t. \quad \textbf{(7)}$$

Inserting this equation into (5) gives recursions for b_t in terms of lags and exogenous variables. These recursions are initialized by exogenously determined b_0 and P_0. We may think of b_0 as representing the initial beliefs of agents and P_0 as yielding a measure of the confidence that agents have in their initial estimates. Like Bray and Savin, our goal is to analyze the process b_t and determine under what conditions it converges to $m/(1-a)$.

We first prove a few results concerning the process P_t. Because it will be used repeatedly, we state the well-known matrix inversion lemma here.

LEMMA 1. *Let W, X, Y, Z be conformable matrices. Then, provided the indicated inverses exist,*

$$[W + XYZ]^{-1} = W^{-1} - W^{-1}X[ZW^{-1}X + Y^{-1}]^{-1}ZW^{-1}.$$

LEMMA 2. *The matrix P_t is symmetric and positive definite provided P_0 is positive definite.*

Proof. This is surely well known, but we provide a proof here for completeness. The proof is by induction. Since $\text{var}(\eta_t)$ is symmetric and positive definite, equation (6) shows that it suffices to prove that

$$P_{t-1} - P_{t-1} x_t D^{-1} x_t' P_{t-1}$$

is symmetric and positive definite, where $D = \sigma^2 + x_t' P_{t-1} x_t > 0$. Symmetry is trivial. By induction, the matrix inversion lemma may be applied with $W = P_{t-1}$, $X = P_{t-1} x_t$, $Y = -D^{-1}$, and $Z = x_t' P_{t-1}$ to show that

https://doi.org/10.1017/S1365100502010325 Published online by Cambridge University Press

$$[P_{t-1} - K_t x_t x_t' P_{t-1}]^{-1} = P_{t-1}^{-1} + \frac{x_t x_t'}{\sigma^2},$$

thus showing that $[P_{t-1} - K_t x_t x_t' P_{t-1}]^{-1}$ is positive definite and the result follows. ∎

LEMMA 3. $[P_t - \text{var}(\eta_t)]^{-1} = P_{t-1}^{-1} + \frac{1}{\sigma^2} x_t x_t'$.

Proof. It is not obvious that $P_t - \text{var}(\eta_t)$ is invertible. However, since P_{t-1} is invertible, Lemma 1 applies to the expression $P_{t-1} - P_{t-1} x_t D^{-1} x_t' P_{t-1}$ and the result follows by induction. ∎

We now transform the Kalman filter recursions so that the proof of Bray and Savin may be applied directly. To this end, set

$$R_t = \frac{\sigma^2}{t}\left[P_{t-1}^{-1} + \frac{1}{\sigma^2} x_t x_t'\right]. \tag{8}$$

Note that, by the preceding lemmas, R_t is symmetric and positive definite.[5]

LEMMA 4. $\frac{1}{t} R_t^{-1} x_t = K_t x_t$.

Proof. This is simply algebra. Notice, by Lemma 3,

$$\frac{1}{t} R_t^{-1} = \frac{1}{\sigma^2}(P_t - \text{var}(\eta_t)) \tag{9}$$

which, by recursion (6), shows that

$$\begin{aligned}\frac{1}{t} R_t^{-1} x_t &= \frac{1}{\sigma^2}\left[P_{t-1} - P_{t-1} x_t D^{-1} x_t' P_{t-1}\right] x_t \\ &= \frac{1}{\sigma^2} D^{-1}\left[P_{t-1} x_t \left[\sigma^2 + x_t' P_{t-1} x_t\right] - P_{t-1} x_t x_t' P_{t-1} x_t\right] \\ &= D^{-1} P_{t-1} x_t = K_t x_t.\end{aligned}$$

∎

This lemma allows us to write equation (5) as

$$b_t = b_{t-1} + \frac{1}{t} R_t^{-1} x_t [y_t - b_{t-1}' x_t]. \tag{10}$$

Notice that if $R_t = (1/t)\sum x_i x_i'$, then equation (10) coincides with the recursive least-squares estimator of the linear model $y_t = \beta' x_t + \epsilon_t$. Recursions for this estimator are given by equations (14) and (15) below.[6] Furthermore, using the substitution $V_t = R_t^{-1}$, and plugging in the true data-generating process (7), we see that recursion (10) is identical to recursion (3), up to the process V_t. As we mentioned previously, the proof of Bray and Savin's main result depended not on the specific process V_t, but only its almost sure convergence to $(Ex_t x_t')^{-1}$. Thus, to show almost sure convergence of the process (10) to $m/(1-a)$, it suffices to show R_t converges to $Ex_t x_t'$ almost surely.

https://doi.org/10.1017/S1365100502010325 Published online by Cambridge University Press

Analysis of R_t is simplified using the following result.

LEMMA 5. *The recursion for R_t may be written*

$$\rho_t(R_{t-1}, \mathrm{var}(\eta_{t-1})) = -\frac{t(t-1)^2}{\sigma^2} R_{t-1} \left[\frac{t-1}{\sigma^2} R_{t-1} + \mathrm{var}(\eta_{t-1})^{-1}\right]^{-1} R_{t-1} \quad \mathbf{(11)}$$

$$R_t = R_{t-1} + \frac{1}{t}(x_t x_t' - R_{t-1}) + \frac{1}{t^2}\rho_t(R_{t-1}, \mathrm{var}(\eta_{t-1})). \qquad \mathbf{(12)}$$

Proof. By (9),

$$P_{t-1}^{-1} = \left[\frac{\sigma^2}{t-1} R_{t-1}^{-1} + \mathrm{var}(\eta_{t-1})\right]^{-1}.$$

Insert this into (8) and apply Lemma 1 to complete the proof. ■

2.2.1. Back to Bray and Savin. Bray and Savin considered the case in which agents believed the parameters of the model to be constant. This is equivalent to the agents in our model believing that $\mathrm{var}(\eta_t) = 0$. However, whereas Bray and Savin's agents used OLS to form their estimates, our agents use the Kalman filter. In this subsection, we show that these estimators are equivalent.[7] This is not difficult. Indeed,

$$\begin{aligned}
&\lim_{\|\mathrm{var}(\eta_{t-1})\|\to 0} (\rho_t(R_{t-1}, \mathrm{var}(\eta_{t-1}))) \\
&= \lim_{\|\mathrm{var}(\eta_{t-1})\|\to 0} \left(\frac{t(t-1)^2}{\sigma^2} R_{t-1}\mathrm{var}(\eta_{t-1}) \left[\frac{t-1}{\sigma^2} R_{t-1}\mathrm{var}(\eta_{t-1}) + I\right]^{-1} R_{t-1}\right) \\
&= 0.
\end{aligned}$$

The recursions defining the Kalman filter estimator, then, reduce to

$$y_t = (m + ab_{t-1})' x_t + \nu_t, \qquad \mathbf{(13)}$$

$$b_t = b_{t-1} + \frac{1}{t} R_t^{-1} x_t [y_t - b_{t-1}' x_t], \qquad \mathbf{(14)}$$

$$R_t = R_{t-1} + \frac{1}{t}(x_t x_t' - R_{t-1}). \qquad \mathbf{(15)}$$

These recursions are the same as those obtained by Bray and Savin and show that their model is a special case of the model we consider here.

2.2.2. The Bullard result. Bullard (1992) showed, for a class of models that includes ours, that if agents believe the parameters of the perceived model to follow a random walk and if the conditional variance of the random walk is constant, then convergence to REE cannot obtain. The idea behind this result is quite simple. For convergence to occur (and to apply the main results of the theory of stochastic

https://doi.org/10.1017/S1365100502010325 Published online by Cambridge University Press

approximation), the gain of the algorithm must go to zero. In the Kalman filter recursions, this gain is represented by

$$K_t = P_{t-1}\left[\sigma^2 + x_t' P_{t-1} x_t\right]^{-1}.$$

For this term to go to zero (almost surely), P_t, as given by

$$P_t = P_{t-1} - K_t x_t x_t' P_{t-1} + \mathrm{var}(\eta_t),$$

must go to zero (almost surely). However, if $\mathrm{var}(\eta_t) = Q > 0$, this cannot happen. Note that P_t represents the agents' perceived mean squared error at time t. If agents believe that the parameters of the model will always have some nonzero constant conditional variance, they will always believe that the MSE is nonzero, and in fact, bounded away from zero by the conditional variance of the parameters. Further, if their perceived MSE is strictly positive and bounded away from zero, agents will always be willing to adjust their estimates in the presence of forecast error, forecast error which will occur because of the stochastic nature of the model. Thus, the agents' estimators cannot possibly converge to a constant value.

Also, we note here that if $\mathrm{var}(\eta_t)$ does not converge to zero, then Bullard's result still holds. To show this, it suffices to show that, in this case, P_t does not converge to zero. Since $\mathrm{var}(\eta_t)$ does not converge to zero, there is a subsequence, indexed, say, by $t(k)$, which is bounded away from zero. Since $P_t \geq \mathrm{var}(\eta_t)$, it follows that $P_{t(k)}$ is bounded away from zero, and thus P_t cannot converge to zero.[8]

2.2.3. Vanishing variance. The results of the preceding section indicate that a necessary condition for convergence to the REE is that the conditional variance of the random walk be decreasing to zero. And, as mentioned in the introduction, we believe that this is a reasonable assumption to make, for if agents initially use OLS to estimate their parameters, then Bray and Savin's result implies that the conditional variance of the parameters does decrease (eventually) since convergence of the agents' estimators to a constant value does occur. In this section, we take as given that agents believe that the conditional variance of the random walk is decreasing to zero and consider what rate is sufficient to guarantee convergence to the REE.

Recall that the recursion describing R_t is given by

$$R_t = R_{t-1} + \frac{1}{t}(x_t x_t' - R_{t-1}) - \frac{1}{t^2}\rho_t(R_{t-1}, \mathrm{var}(\eta_{t-1})). \qquad \textbf{(16)}$$

Further, we have seen that the form of the Kalman filter recursions, properly transformed, together with the proof of Bray and Savin's main result, shows that if $a < 1$, then convergence of b_t to $m/(1-a)$ occurs with probability 1, provided R_t converges to $E x_t x_t'$ almost surely. Analysis of stochastic processes of the form (16) may be done using the theory of stochastic approximation. Under certain restrictions on the functions ρ_t, a differential equation can be analyzed to determine

https://doi.org/10.1017/S1365100502010325 Published online by Cambridge University Press

possible points of convergence. Unfortunately, because of the form of ρ_t, only local convergence results can be applied and the restrictions on the conditional variance are strong. This is discussed in detail in Section 3. Fortunately, it is possible to prove global convergence (in the sense described later) with weaker restrictions, using a less direct approach. Specifically, we obtain the following result.

LEMMA 6. *If R_1 is positive definite and symmetric, and*

$$\limsup_t t^2 \|\mathrm{var}(\eta_t)\| = 0, \tag{17}$$

then R_t converges to $Ex_t x_t'$ almost surely.

The proof of this lemma is in Appendix A. The implication of restriction (17) is that the norm of the conditional variance must die just a little faster than $1/t^2$. Also, the convergence is global with respect to the initial condition, subject to the restriction that the initial condition is symmetric, positive definite. Note that R_1 is symmetric, positive definite provided P_0 is, and that P_0 represents the agents' perceived mean squared error of their initial belief, b_0. We conclude that this is not a significant restriction.

Lemma 6, together with the previous observations concerning the application of Bray and Savin's proof to our recursions, yields the following theorem which is the main result of this paper.

THEOREM 2. *If P_0 is positive definite and*

$$\limsup_t t^2 \|\mathrm{var}(\eta_t)\| = 0,$$

then $b_t \to m/(1-a)$ almost surely, provided that $a < 1$.

Observe that the restriction on the parameter a, called the stability parameter, is the same as obtained by Bray and Savin. This is not surprising. The recursions describing the time path of b_t are the same for both models, except for the value of the positive definite matrix modifying the forecast error. Specifically, both recursions can be written as

$$b_t = b_{t-1} + \frac{1}{t} V_t x_t ((1-a) b_{t-1}' x_t + m' x_t + \nu_t),$$

and the only difference will be the values of the positive definite matrix V_t. The restriction on a guarantees that b_t moves toward its REE value. Because V_t is positive definite in both models, its specific value does not affect this direction.

The theorem predicts convergence provided that agents believe that the conditional variance of the random walk eventually decreases a little faster than t^{-2}. Also, as Bullard showed, convergence is not obtained if the conditional variance is constant. What happens when the conditional variance decreases to zero at a rate less than or equal to t^{-2} has not been determined analytically. In a companion paper (McGough, Forthcoming), we report the results of simulations indicating that convergence can be obtained for rates of decrease slightly less than t^{-2} and also that for slow rates of decrease, say $t^{-0.5}$, convergence does not appear to occur.

3. A MORE GENERAL COBWEB MODEL

The analysis of stochastic recursive algorithms is usually done using the theory of stochastic approximation. For economic models, the standard is to use the results of Ljung (1977), Marcet and Sargent (1989), and Evans and Honkapohja (2001), which tell us to associate a differential equation with the given stochastic process. It can then be shown that, under certain conditions, possible convergence points of the process correspond to stable fixed points of the differential equation.[9] In this section, we apply the theory of stochastic approximation to the sequence of estimators obtained from a more general cobweb model. We obtain local convergence results, provided that stronger restrictions are placed on the rate of decrease of the conditional variance.

3.1. Model

We generalize the model used by Bray and Savin to include serially correlated, observable shocks. Specifically, we consider a reduced-form Muth model as given by

$$\begin{aligned} y_t &= aE^*_{t-1}y_t + \lambda' x_t \\ x_t &= Bx_{t-1} + \nu_t, \end{aligned} \tag{18}$$

where $x_t \in R^n$ is an asymptotically stationary process with the first component equal to 1 and $y_t \in R$. For technical reasons, we require the i.i.d. process ν_t to be almost surely uniformly bounded. Deviating slightly from the assumptions of the previous model, we assume that the expected time-t value of the endogenous variable is formed with respect to information available at time $t-1$. This information includes x_{t-1}. This form of the model has been studied by, for example, Evans and Honkapohja (2001).

The unique REE for this model is given by

$$y_t = \frac{\lambda' B}{1-a}x_{t-1} + \lambda'\nu_t.$$

To incorporate learning into the model, we assume that agents believe the final form of the model to be

$$\begin{aligned} y_t &= \beta_t' x_{t-1} + \varepsilon_t, \\ \beta_{t+1} &= \beta_t + \eta_t. \end{aligned} \tag{19}$$

Notice that this is the same form of beliefs as in the previous model, except for the timing of the exogenous variable.

https://doi.org/10.1017/S1365100502010325 Published online by Cambridge University Press

Given the beliefs of the agents, the natural estimator is again given by the Kalman filter. Denote by b_t the Kalman filter estimate of β_{t+1} using information available at time t. Then, at time $t-1$, agents believe that y_t will be determined by the following equation, called the perceived law of motion, or PLM:

$$y_t = b'_{t-1} x_{t-1} + \varepsilon_t. \tag{20}$$

Using the PLM, agents determine $E^*_{t-1}(y_t) = b'_{t-1} x_{t-1}$. This may be inserted into equation (18) to obtain the actual data-generating process. Define the map $T : R^n \to R^n$ by

$$T(b) = ab + \lambda' B. \tag{21}$$

Then, the data are generated by the following equation, called the actual law of motion, or ALM:

$$y_t = T(b_{t-1})' x_{t-1} + \lambda' \nu_t. \tag{22}$$

The map T takes the perceived parameters to the actual parameters. When the perceived parameters equal the actual parameters, the model in is an REE; thus, a fixed point of the $T-$map determines an REE.

The recursions describing the agents' estimators can now be reported. Since the state-space model (19) describing agents beliefs is identical, up to timing, to the state-space model considered earlier, the recursions describing the Kalman filter estimator are identical up to timing as well. We obtain

$$K_t = P_{t-1}\left[\sigma^2 + x'_{t-1} P_{t-1} x_{t-1}\right]^{-1}, \tag{23}$$

$$b_t = b_{t-1} + K_t x_{t-1}[(T(b_{t-1}) - b_{t-1})' x_{t-1} + \lambda' \nu_t], \tag{24}$$

$$P_t = P_{t-1} - K_t x_{t-1} x'_{t-1} P_{t-1} + \mathrm{var}(\eta_t), \tag{25}$$

where the actual law of motion has been inserted for y_t. As before, our goal is to analyze the asymptotic behavior of b_t.

3.2. Convergence

To analyze the asymptotic behavior of the agents' estimators, we use the theory developed by Evans and Honkapohja (2001); for a summary, see Evans and Honkapohja (1998). Consider a recursive stochastic algorithm of the following form:

$$\begin{aligned} \theta_t &= \theta_{t-1} + \frac{1}{t} H(\theta_{t-1}, w_t) + \frac{1}{t^2} \rho_t(\theta_{t-1}, w_t), \\ w_t &= A(\theta_{t-1}) w_{t-1} + B(\theta_{t-1}) \mu_t, \end{aligned} \tag{26}$$

where μ_t is white noise. To this process is associated a differential equation as follows. Set

$$h(\theta) = \lim_{t \to \infty} E(H(\theta, x_t)). \tag{27}$$

https://doi.org/10.1017/S1365100502010325 Published online by Cambridge University Press

The differential equation is

$$\frac{d\theta}{dt} = h(\theta). \tag{28}$$

The main result of the theory says that if θ^* is a locally asymptotically stable fixed point of this differential equation, and H and ρ_t satisfy some nice properties in some neighborhood of this fixed point, then the process (26) converges to θ^* with probability 1, provided a projection facility is used; see Appendix B for details.[10]

To employ the theory described above, we must put our algorithm into the form (26). This is done using the same variable substitution that was used to analyze the previous model. Specifically, set

$$R_t = \frac{\sigma^2}{t}\left[P_{t-1}^{-1} + \frac{1}{\sigma^2}x_{t-1}x_{t-1}'\right].$$

Then, the Kalman filter recursions can be rewritten as

$$\hat{\rho}_t(b_{t-1}, R_{t-1}) = -\frac{t(t-1)^2}{\sigma^2}R_{t-1}Q_{t-1}\left[\frac{t-1}{\sigma^2}R_{t-1}Q_{t-1} + I\right]^{-1}R_{t-1}, \tag{29}$$

$$R_t = R_{t-1} + \frac{1}{t}(x_{t-1}x_{t-1}' - R_{t-1}) + \frac{1}{t^2}\hat{\rho}_t(b_{t-1}, R_{t-1}), \tag{30}$$

$$b_t = b_{t-1} + \frac{1}{t}R_t^{-1}x_{t-1}((T(b_{t-1}) - b_{t-1})'x_{t-1} + \lambda'\nu_t), \tag{31}$$

where $Q_t = \text{var}(\eta_t)$. The dependence of b_t on R_t forces us to make the standard variable change $S_{t-1} = R_t$. Letting $\theta_t = [b_t', S_t']'$ then allows us to write the Kalman filter recursions in the form (26),[11] where $w_t = [x_t', x_{t-1}', \nu_t]'$,

$$H(\theta_{t-1}, w_t) = \begin{bmatrix} S_{t-1}^{-1}x_{t-1}((T(b_{t-1}) - b_{t-1})'x_{t-1} + \lambda'\nu_t) \\ (x_t x_t' - S_{t-1}) \end{bmatrix}, \tag{32}$$

$$\rho_t(\theta_{t-1}, w_t) = \begin{bmatrix} 0 \\ -\left(\frac{t}{t+1}\right)\left[\frac{t^3}{\sigma^2}S_{t-1}Q_t\left[\frac{t}{\sigma^2}S_{t-1}Q_t + I\right]^{-1}S_{t-1} + x_t x_t' - S_{t-1}\right] \end{bmatrix}. \tag{33}$$

Having placed our stochastic algorithm in the correct form, we now consider its associated differential equation. It is easily computed to be

$$\frac{db}{d\tau} = S^{-1}M(T(b) - b), \tag{34}$$

$$\frac{dR}{d\tau} = M - S, \tag{35}$$

where $M = \lim_{t\to\infty} Ex_t x_t'$ exists and is positive definite because x_t is asymptotically stationary.

https://doi.org/10.1017/S1365100502010325 Published online by Cambridge University Press

Just as in the previous model, the Kalman filter estimator reduces to the OLS estimator as $\|Q_t\| \to 0$. Since Q_t is only present in the ρ_t term, it does not affect the functional form of the associated differential equation. Thus, we are not surprised to find that the differential equation above is identical to the one obtained when least-squares learning is modeled. Also, as has been shown in the literature on least-squares learning, the unique fixed point of the preceding system is $b^* = (\lambda' B)/(1-a)$ and $S = M$, and this fixed point is locally asymptotically stable provided that $a < 1$.

According to the preceding paragraph, to show convergence to REE, it suffices to restrict the model so that the technical conditions on H and ρ_t are satisfied. As mentioned, the form of H seen here is not new; it is identical to the H obtained from least-squares learning models. In particular, that it satisfies the appropriate conditions is well known: see, for example, Evans and Honkapohja (2001). On the other hand, the form of ρ_t is new and thus the restrictions must be considered.

Because the technical nature of the formal restrictions on ρ_t would be distracting, we will work informally here and relegate the formal details to Appendix B. Intuitively, provided ρ_t does not get big, then, when scaled by t^{-2}, it becomes irrelevant to the asymptotic behavior of R_t. More specifically, if ρ_t is bounded in t the theorem of Evans and Honkapohja may be applied.[12] For ρ_t to be bounded in t, it is clear, given equation (33), that the conditional variance, Q_t, must vanish like t^{-3}. To this end, we make the following assumption:

$$\sup_{t>0} t^3 \|Q_t\| = \sigma^2 k < \infty. \tag{36}$$

We have following theorem.

THEOREM 3. *If the economy is given by the model* (18) *and if agents have beliefs given by* (19) *together with the restriction on the conditional variance*, (36), *and if agents use the Kalman filter to obtain their expectations*, *and if the value of the stability parameter a is less than* 1, *then the economy will converge to the REE with probability* 1 *provided the learning algorithm is augmented with a projection facility.*

The proof of this result is contained in Appendix B.

3.3. E-Stability

In this section, we make an important link to the learning literature. To determine the stability under learning of a given REE, the industry standard is to use the theory of E-stability.[13] This theory tells us to consider the differential equation[14]

$$\frac{db}{d\tau} = T(b) - b. \tag{37}$$

Notice that an REE, b^*, is a stationary solution to this differential equation. The REE is said to be *E-stable* provided that it is locally asymptotically stable. The

E-stability principle says that an E-stable REE is locally stable under learning provided that a reasonable learning algorithm is employed. The intuition behind this principle is not difficult. Suppose that, by reasonable learning algorithm, it is meant that the new parameter estimates are obtained by moving in the direction of the forecast error redirected appropriately by the value of the regressor.[15] Given the actual law of motion, this product may be written

$$x_{t-1}x'_{t-1}(T(b_{t-1}) - b_{t-1}) + x_{t-1}\nu'_t\lambda.$$

Because $x_{t-1}x'_{t-1}$ is non-negative definite, the components of the vector determined by the preceding expression will, on average, have the same signs as the components of $T(b_{t-1}) - b_{t-1}$. Thus, if the differential equation above is locally asymptotically stable at b^*, then moving according to the learning algorithm should, on average, result in convergence to b^*. Note that the E-stability principle is not a general result, and so, whenever possible, convergence should be proven using other techniques.

Recall that a differential equation of the form (37) is locally asymptotically stable at b^* provided the eigenvalues of the derivative have real part less than zero. In our case, the derivative is $DT - I$, and so, local asymptotic stability requires that the eigenvalues of DT must have real part less than 1. Since $DT = aI$, it follows that the REE is E-stable provided $a < 1$.[16] It is well known that $a < 1$ implies convergence to REE of the least-squares learning algorithm: See, for example, Evans and Honkapohja (2001). Thus, E-stability determines convergence when least-squares learning is used. We have shown that E-stability also determines convergence when Kalman filter learning is used, provided the conditional variance of the random walk decreases rapidly.

4. CONCLUSION

Since its conception, the assumption of rational agents has been criticized as being too strong. The landmark paper of Bray and Savin, and the learning literature published since, has given credence to the rational expectations hypothesis because it has shown that, for many models, weakening the assumption to that of boundedly rational agents preserves rational expectations equilibria asymptotically. However, these least-squares learning models have been plagued by the same criticism originally borne by the simple adaptive models that predated and, in fact, led to the rational expectations hypothesis: Why would reasonable agents make systematic errors?[17]

In this paper we have begun the process of addressing the issue of model misspecification. We have allowed our agents to increase their sophistication by postulating a time-varying process for the parameters of the model. We have shown that, for certain restrictions on the postulated process, convergence to REE still obtains. This further strengthens the learning literature's justification for continued analysis of REE. And these results are in contrast to the nonconvergence result obtained by Jim Bullard. He shows that if the conditional variance of the random walk is constant, then convergence cannot possibly obtain. We have argued that it is more natural to consider the specification that the conditional variance decreases to zero and subsequently overturned his result.

The results of this paper are not complete. Several important questions remain. First, what is the asymptotic behavior of the economy if the conditional variance of the random walk decreases to zero more slowly than the restrictions required for our results? Simulations that are reported in a companion paper suggest that convergence is obtained for rates of decrease that are slower, but not all such rates. Second, what is the effect of altering the assumption that the parameters follow a random walk? Bray and Savin suggest a return-to-normalcy process; such a process with constant conditional variance is analyzed by Bullard and he again obtains a nonconvergence result. Analysis of such a model with decreasing variance has proven difficult because the form of the resulting recursions is not addressed by the stochastic approximation literature. Simulations reported in the companion paper suggest that convergence to REE is obtained, provided the conditional variance decreases rapidly enough and that the restrictions on this rate of decrease may be weaker than the analogous restrictions in the random-walk model. Finally, our agents, just like Bray and Savin's, have a misspecified model. Can the agents use some natural econometric technique to detect this misspecification? We are working on this problem currently.

NOTES

1. This conclusion is explained more fully in Section 2.
2. Margaritis (1990) also considers a model with time-varying parameters. For convergence to a point to occur, his results require that the gain of the adaptive algorithm tends to zero. As we note in Section 2.2.2, and as was shown by Bullard (1992), this cannot hold in our model if the conditional variance of the time-varying parameters is positive definite.
3. For a careful derivation of this reduced form, see Bray and Savin (1986).
4. Agents are free to use x_t in their estimation of b_{t-1} but the independence assumption made above implies that the realization of x_t will not alter their estimates.
5. This variable transform was used by Margaritis (1990).
6. For more on the RLS algorithm, see Evans and Honkapohja (2001).
7. This type of equivalence was noted in a different model by Bullard (1992).
8. The notion of "greater than" for matrices is reviewed in Appendix A.
9. Note the power of this type of result. It tells us not only that convergence occurs, but also yields the possible limit points. And, being fixed points of differential equations, these limit points are often not difficult to compute.
10. Informally, a projection facility puts the process back near the fixed point if it wanders too far away. For details, see Appendix B. Projection facilities were introduced by Ljung (1977) and Marcet and Sargent (1989). Evans and Honkapohja (1998, 2001) also show how weaker results can be obtained if the projection facility is dropped.
11. We are being a little sloppy here. Technically, the theorems apply to vector processes $\boldsymbol{\theta}$, but in our case, R is a matrix. The complication is avoided via application of the column operator, which takes a matrix to the associated column vector. We suppress this for notational simplicity as is standard in the literature.
12. Again, additional restrictions apply; see Appendix B for details.

13. Originally coined by George Evans.

14. The $T-$map, and hence E-stability, can be defined for many different types of models. Also, the $T-$map depends only on the reduced-form equation and the agents' perceived beliefs and thus is independent of parameter estimation procedure.

15. Both the OLS estimator and the Kalman filter estimator behave in this manner.

16. Of course, this is the same restriction on the stability parameter as obtained by Bray and Savin for their slightly different model.

17. The error in the models of least-squares learning—the misspecification of the model—is, admittedly, more subtle than before, but still sometimes detectable using standard econometric techniques.

18. By "surely well known," I mean that it is easier to prove than to try to look up, but I deserve no credit for its discovery.

REFERENCES

Bray, M. & N. Savin (1986) Rational expectations, learning, and model specification. *Econometrica* 57, 1129–1160.

Brockwell, P. & R. Davis (1987) *Time Series: Theory and Methods*. Springer-Verlag.

Bullard, J. (1992) Time-varying parameters and non-convergence to rational expectations equilibrium under least squares learning. *Economics Letters* 40, 159–166.

Evans, G. & S. Honkapohja (1998) Learning dynamics. In J. Taylor & M. Woodford (eds.), *Handbook of Macroeconomics*. Amsterdam: Elsevier.

Evans, G. & S. Honkapohja (2000) *Learning and Expectations in Macroeconomics*. Princeton, NJ: Princeton University Press.

Kadison, R. & J. Ringrose (1983) *Fundamentals of the Theory of Operator Algebras*, vol. 1. Academic Press.

Ljung, L. (1977) Analysis of stochastic recursive algorithms. *IEEE Transactions on Automatic Control* 22, 551–575.

Marcet, A. & T. Sargent (1989) Convergence of least squares mechanisms in self-referential linear stochastic models. *Journal of Economic Theory* 48, 337–368.

Margaritis, D. (1990) A time-varying model of rational learning. *Economics Letters* 33, 309–314.

McGough, B. (Forthcoming) A Numerical Study of Learning Models with Time-Varying Parameters.

Stokey, N. & R. Lucas, with E. Prescott (1989) *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard University Press.

APPENDIX A: CONVERGENCE OF R_t

The argument that follows will repeatedly use the probabilistic notion of *event*. We employ the following notation: Let x be a random variable and P be some property that realizations of x may or may not have. Then, $E = \{x \text{ has } P\}$ means E is the event that the realization of x has the property P. Further, if E and F are events, then $E \subset F$ means the event F occurs whenever the event E occurs. Also, for a given random variable w, denote by $\tilde{w}$ a particular realization. Notice that if $E = \{x \text{ has } P\}$ and $\tilde{x}$ has P, then E has occurred.

The following results concern almost sure convergence of stochastic processes. Recall that the process x_n converges to x almost surely (that is, with probability 1) provided the event that the sequence of realizations of x_n converges to the realization of x occurs with probability 1. In the sequel, we make repeated use of the following lemma.

LEMMA A.1. *The sequence of random variables x_n converges almost surely to the random variable x if and only if for any $\epsilon > 0$ and any $p \in (0, 1)$ there is an N so that*

$$Pr\{\|x_n - x\| \leq \epsilon \, \forall n \geq N\} > p.$$

This result follows immediately from Lemma 14.1 in Stokey and Lucas (1989).

We say a matrix is positive if it is symmetric and positive definite, and nonnegative if it is symmetric and positive semidefinite. If A and B are symmetric, we say $A < B$ if $B - A$ is positive (or written $B - A > 0$) and we say $A \leq B$ if $B - A$ is nonnegative (or written $B - A \geq 0$.) These relations induce partial orderings on the set of symmetric matrices. It is well known that a matrix is positive (nonnegative) provided all eigenvalues are positive (nonnegative and real).

The recursion R_t is defined as follows: Let $x_t \in R^n$ be i.i.d. Then,

$$\rho_t(R_{t-1}, Q_{t-1}) = \frac{(t-1)^2}{\sigma^2} R_{t-1} Q_{t-1} \left[\frac{t-1}{\sigma^2} R_{t-1} Q_{t-1} + I\right]^{-1} R_{t-1},$$

$$R_t = R_{t-1} + \frac{1}{t}(x_t x_t' - R_{t-1}) - \frac{1}{t}\rho_t(R_{t-1}, Q_{t-1}).$$

We assume that Q_t is positive for all t. We take the initial condition of the recursion to be a positive matrix and recall the previously mentioned implication that, with probability 1, all elements of the sequence R_t will be positive. Assume $\limsup_t t^2 \|Q_t\| = 0$ and set $E(x_t x_t') = \Omega$. Our goal is to prove the following result.

THEOREM A.1. *If R_1 is positive, then the process R_t converges to Ω almost surely.*

This will be facilitated by the following lemma on positive matrices and the induced partial ordering.

LEMMA A.2. *Let A, B, and C be positive conformable matrices.*

(i) *If $A \leq B$ and if C commutes with A and B then $AC \leq BC$.*
(ii) *If $A \leq B$, then $B^{-1} \leq A^{-1}$.*
(iii) *$A \leq \|A\| I$.*
(iv) *If $A \leq B$, then $\|A\| \leq \|B\|$.*

These results are standard. The first three can be found in Kadison and Ringrose (1983). To prove the fourth, proceed as follows: Statement (iii) implies $A \leq \|B\| I$. Note that λ is an eigenvalue of A if and only if $\|B\| - \lambda$ is an eigenvalue of $\|B\| I - A$. Also, since A is symmetric, $\|A\| = \max_i \|\lambda_i\|$ where λ_i varies over the eigenvalues of A. Let λ_m be the eigenvalue of maximum modulus. Note that it is also real and positive. Since $\|B\| I - A \geq 0$, its eigenvalues must be nonnegative. Thus, $\|B\| - \|A\| = \|B\| - \lambda_m \geq 0$.

The following result relating the partial ordering on symmetric matrices to convergence is needed.

LEMMA A.3. *Suppose x_n is a sequence of stochastic matrices that are almost everywhere positive and that converge almost surely to the positive matrix x. Fix some $\delta > 0$ and, for given N, define the event E as*

$$E(N) = \{x_n \leq (\|x\| + \delta) I \;\forall\; n \geq N\}.$$

Then, for any $p \in (0, 1)$, there exists N so that $Pr\{E(N)\} > p$.

Proof. Define the event F as follows:

$$F(N) = \{\|x_n\| \leq \|x\| + \delta \ \forall \ n \geq N\}.$$

Because $x_n \to x$ a.s., we may choose N so that $\Pr\{F(N)\} > p$. Now notice that whenever $F(N)$ occurs,

$$x_n \leq \|x_n\| I \leq (\|x\| + \delta) I, \ \forall n \geq N,$$

which implies $F(N) \subset E(N)$. ■

The next lemma is used to provide an upper bound for the sequence R_t. Define a new sequence as follows:

$$\hat{R}_t = \hat{R}_{t-1} + \frac{1}{t}(x_t x_t' - \hat{R}_{t-1}).$$

Note that $\hat{R}_t = (1/t) \sum_{i=1}^{t} x_i x_i'$ so that, by the law of large numbers, $\hat{R}_t$ converges to Ω almost surely.

LEMMA A.4. $R_t \leq \hat{R}_t$.

Proof. The proof is by induction. Let $\hat{R}_1 = R_1$ and assume $R_{t-1} \leq \hat{R}_{t-1}$. Also, notice $\rho_t(R_{t-1}, Q_{t-1})$ is positive. Then,

$$\begin{aligned} R_t &= R_{t-1} + \frac{1}{t}(x_t x_t' - R_{t-1}) - \frac{1}{t}\rho_t(R_{t-1}, Q_{t-1}) \\ &= \frac{t-1}{t} R_{t-1} + \frac{1}{t} x_t x_t' - \frac{1}{t}\rho_t(R_{t-1}, Q_{t-1}) \\ &\leq \frac{t-1}{t} \hat{R}_{t-1} + \frac{1}{t} x_t x_t' = \hat{R}_t. \end{aligned}$$

■

Now that we have established an upper bound for R_t, we begin work on the lower bound. Fix $\delta > 0$ and let $\xi = \|\Omega\| + \delta$, and $M(T) = \sup_{t>T}(t^2/\sigma^2)\|Q_t\|$. Notice that, by assumption, $M(T) \to 0$ and is decreasing. For each T, we define a new sequence as follows:

$$S_t(T) = \begin{cases} R_t & \text{if } t \leq T \\ S_{t-1}(T) + \frac{1}{t}(x_t x_t' - S_{t-1}(T)) - \frac{1}{t} M(T)\xi^2 I & \text{else.} \end{cases}$$

It is simple to show that $S_t(T)$ converges to $\Omega - M(T)\xi^2 I$ almost surely. Indeed, a straightforward induction argument shows $S_t(T) = (1/t)\sum_{i=1}^{t} x_i x_i' - M(T)\xi^2 I$, and the law of large numbers completes the result. The following lemma relates these sequences to R_t.

LEMMA A.5. *Let* $p \in (0, 1)$ *and*

$$E(T) = \{S_t(T) \leq R_t \, \forall t\}.$$

Then there exists a $\hat{T}$ *so that* $T \geq \hat{T}$ *implies* $Pr(E(T)) > p$.

Proof. Set

$$F(T) = \{\hat{R}_t \leq \xi I \, \forall t \geq T\}.$$

By Lemma A.3, there is a $\hat{T}$ so that $\Pr(F(\hat{T})) > p$. To complete the proof then, it suffices to show that $T_2 \geq T_1 \Rightarrow F(T_1) \subset E(T_2)$. The proof is by induction. Assume $F(T_1)$ occurs.

https://doi.org/10.1017/S1365100502010325 Published online by Cambridge University Press

By construction, $S_t(T_2) \leq R_t$ for $t \leq T_2$. So, let $t > T_2$ and assume $S_{t-1}(T_2) \leq R_{t-1}$. Since $F(T_1)$ occurred and $t > T_2 > T_1$, it follows, using Lemma A.2, that

$$\begin{aligned}\rho_t(R_{t-1}, Q_{t-1}) &= \frac{(t-1)^2}{\sigma^2} R_{t-1} Q_{t-1} \left[\frac{t-1}{\sigma^2} R_{t-1} Q_{t-1} + I\right]^{-1} R_{t-1} \\ &\leq M(T_2) R_{t-1} \left[\frac{t-1}{\sigma^2} R_{t-1} Q_{t-1} + I\right]^{-1} R_{t-1} \\ &\leq M(T_2) R_{t-1}^2 \leq M(T_2)\xi^2 I.\end{aligned}$$

It follows that

$$\begin{aligned}S_t(T_2) &= S_{t-1}(T_2) + \frac{1}{t}(x_t x_t' - S_{t-1}(T_2)) - \frac{1}{t} M(T)\xi^2 I \\ &= \frac{t-1}{t} S_{t-1}(T_2) + \frac{1}{t} x_t x_t' - \frac{1}{t} M(T)\xi^2 I \\ &\leq \frac{t-1}{t} R_{t-1} + \frac{1}{t} x_t x_t' - \frac{1}{t} M(T)\xi^2 I \\ &\leq \frac{t-1}{t} R_{t-1} + \frac{1}{t} x_t x_t' - \frac{1}{t} \rho_t(R_{t-1}, Q_{t-1}) = R_t.\end{aligned}$$

■

We are now ready to prove the main result.

Proof. Let $p \in (0, 1)$ and $\epsilon > 0$. As usual, we have

$$\|R_t - \Omega\| \leq \|R_t - S_t(T)\| + \|S_t(T) - \hat{R}_t\| + \|\hat{R}_t - \Omega\|.$$

Set

$$D(K) = \left\{ \|\hat{R}_t - \Omega\| < \frac{\epsilon}{3} \ \forall\ t \geq K \right\}$$

and choose $\tilde{K}$ so that $\Pr(D(\tilde{K})) > (2+p)/3$. Choose T_1 so that $M(T_1)\xi^2 < \epsilon/6$ and choose $T_2 > T_1$ so that $\Pr(E(T_2)) > (2+p)/3$, where $E(T)$ is the event as defined in Lemma A.5. This fixes the sequence $S_t(T_2)$. Define the event F by

$$F = \{0 \leq \|R_t - S_t(T_2)\| \leq \|\hat{R}_t - S_t(T_2)\| \ \forall\ t\}.$$

Notice that, by statement (iv) of Lemma A.2, $E(T_2) \subset F$.

Since $S_t(T_2) \to \Omega - M(T_2)\xi^2 I$ and $\hat{R}_t \to \Omega$ almost surely, it follows that $\|\hat{R}_t - S_t(T_2)\| \to M(T_2)\xi^2$ almost surely. Let

$$G(K) = \left\{ \|\hat{R}_t - S_t(T_2)\| < M(T_2)\xi^2 + \frac{\epsilon}{6} \ \forall\ t \geq K \right\}.$$

Then we may choose $\hat{K} > \tilde{K}$ so that $\Pr(G(\hat{K})) > (2+p)/3$.

Now, let $H(K)$ be the following event:

$$H(K) = \{\|R_t - \Omega\| < \epsilon \ \forall\ t \geq K\}.$$

https://doi.org/10.1017/S1365100502010325 Published online by Cambridge University Press

The main result is proved by showing that we may choose K so that $\Pr(H(K)) > p$. We claim that $\hat{K}$ suffices. To see this, first notice $G(\hat{K}) \cap F \cap D(\tilde{K}) \subset H(\hat{K})$. Indeed, suppose $G(\hat{K}) \cap F \cap D(\tilde{K})$ occurs. Then, for all $t > \hat{K}$,

$$\begin{aligned}\|R_t - \Omega\| &\le \|R_t - S_t(T_2)\| + \|S_t(T_2) - \hat{R}_t\| + \|\hat{R}_t - \Omega\| \\ &\le M(T_2)\xi^2 + \frac{\epsilon}{6} + M(T_2)\xi^2 + \frac{\epsilon}{6} + \frac{\epsilon}{3} < \epsilon.\end{aligned}$$

Finally, note that

$$\begin{aligned}\Pr(G(\hat{K}) \cap F \cap D(\tilde{K})) &= 1 - \Pr((G(\hat{K}) \cap F \cap D(\tilde{K}))^c) \\ &= 1 - \Pr(G(\hat{K})^c \cup F^c \cup D(\tilde{K})^c) \\ &\ge 1 - [\Pr(G(\hat{K})^c) + \Pr(F^c) + \Pr(D(\tilde{K})^c)] \\ &\ge 1 - 3\left[1 - \frac{2+p}{3}\right] = p. \qquad \blacksquare\end{aligned}$$

APPENDIX B: PROOF OF THEOREM 3

The main result of Section 3 is proved using a theorem by Evans and Honkapohja. We present the remaining details of that theorem's use here. We placed our algorithm in the required form, repeated here for convenience:

$$\theta_t = \theta_{t-1} + \frac{1}{t}H(\theta_{t-1}, w_t) + \frac{1}{t^2}\rho_t(\theta_{t-1}, w_t).$$

To apply the results of Evans and Honkapohja, it must be shown that H and ρ_t satisfy certain conditions. These conditions are carefully reported by Evans and Honkapohja (2001). The form of H in our model is not new; it is well known, and shown by Evans and Honkapohja (2001), that H has the appropriate properties. We proceed to examine the necessary restrictions on ρ_t. The result of Evans and Honkapohja requires ρ_t to be bounded in t by a simple function of w for all S in compact sets surrounding the fixed point. Specifically, we must show there is a U with $M \in U$ so that, given compact K in U, there are constants C and q so that for all t and for all $S \in K$,

$$|\rho_t(S, w)| \le C(1 + |w|^q).$$

It is easier to work with matrix norms. The following lemma, which is surely well known, allows us to do that.[18]

LEMMA B.1. *Let* $A \in R^{n\times n}$, $x_n \in R^n$, *and* $\|A\| = sup_{|v|\le 1}|Av|$ *be the usual matrix norm. Then*,

(i) $n\|A\|^2 \ge (col(A))^2$,
(ii) $\|xx'\| \le n^2|x|^2$.

Proof. (i) It can be shown that $\|A\| = \max_i \sum_j |a_{ij}|$. Thus,

$$n\|A\|^2 = n \max_i \sum_j |a_{ij}|^2 + \sum_{j \neq k} |a_{ij} a_{ik}|$$
$$\geq n \max_i \sum_j |a_{ij}|^2 \geq \sum_{ij} |a_{ij}|^2 = (col(A))^2.$$

(ii)

$$\|xx'\| = \max_i \sum_j |x_i x_j| \leq \sum_{ij} |x_i x_j|$$
$$\leq n^2 \max_{ij} |x_i x_j| = n^2 \max_i x_i^2 \leq n^2 |x|^2. \qquad \blacksquare$$

Recall the restriction on the condition variance Q_t:

$$\sup_{t>0} t^3 \|Q_t\| = \sigma^2 k < \infty. \tag{B.1}$$

Using the fact that $tQ_t/\sigma^2 \to 0$, it is straightforward to show that there exists an open set U about M so that for $S \in U$, the matrix $tQ_t/\sigma^2 S + I$ is invertible, and further, for any compact subset K of U, there exists a positive number N so that $\|(tQ_t/\sigma^2 S + I)^{-1}\| < N$ for all $S \in K$ and for all t. Now let

$$C_1 = \max_{S \in K}\{\|S\|^2 N + \|S\|\},$$
$$C = \sqrt{n} \max\{kC_1, n^2\}.$$

Then

$$|\rho_t(S, x)| \leq \sqrt{n}\|\rho_t(S, x)\| \leq \sqrt{n}(k\|S\|^2 N + \|xx'\| + \|S\|)$$
$$\leq \sqrt{n}\left(C_1 + n^2|x|^2\right) \leq C(1 + |x|^2).$$

We may now proceed to state and apply the result of Evans and Honkapohja. The differential equation (34) is locally asymptotically stable at (b^*, M) provided $a < 1$. Let D be an open domain of attraction of (b^*, M). By the converse to Lyapunov's theorem [see Evans and Honkapohja (2001, Proposition 5.9)], there is a Lyapunov function $U : D \to$ R with the following properties:

(i) $U(b^*, M) = 0$ and $U(\theta) > 0$ for all $\theta \neq (b^*, M)$
(ii) $U(\theta) \to \infty$ as $\theta \to \partial D$.

For $c > 0$, set $K(c) = \{\theta : U(\theta) < c\}$. Note that $c_1 \leq c_2 \Rightarrow K(c_1) \subset K(c_2) \subset D$. For fixed c_2 and $c_1 < c_2$, define the projection facility as follows:

$$P(\theta) = \begin{cases} \theta & \text{if} \quad \theta \in K(c_2) \\ \in K(c_1) & \text{if} \quad \theta \notin K(c_2). \end{cases}$$

The theorem of Evans and Honkapohja allows us to conclude that if the algorithm is augmented with the preceding projection facility, then the process $\theta_t = [\begin{smallmatrix} b_t \\ S_t \end{smallmatrix}]$ converges almost surely to (b^*, M).

https://doi.org/10.1017/S1365100502010325 Published online by Cambridge University Press