A Numerical Approach to Learning with Time-Varying Parameters*

Bruce McGough
Oregon State University

April 29, 2002

Abstract

Bray and Savin [1] raised the concern that even with sophisticated adaptive expectations, agents may still detect the presence of systematic forecast errors, and consequently suggested analyzing models with time-varying parameters perceptions. McGough [6] showed analytically that for certain types of time-varying models, convergence to the rational expectations equilibrium still obtained. We continue this research by analyzing numerically model specifications not addressed in previous papers. We assume the parameters in agents' perceived law of motion (PLM) follows either a random walk or a stationary AR(1) process. We find that if the conditional variance of the PLM tends to zero like $1/t^e$ then convergence to REE appears to obtain for e as low as 1.09 in the random walk model and for e as low as .52 in the AR(1) model. These values are considerably lower than the sufficient bound of $e > 2$ obtained analytically in [6] for the random walk model, and support the assertion that even in case agents recognize misspecification in their model, convergence to REE may still obtain.

1 Introduction

In their seminal work, [1], Bray and Savin used a simple cobweb model to analyze the asymptotics of an economy whose agents use ordinary least squares to update their misspecified beliefs. They found that for some parameter values, convergence to the rational expectations equilibrium obtains. The misspecification in Bray and Savin's model arrises because while agents believe the parameters of the model are time-invariant, these parameters are in fact endogenously determined and vary with time. This misspecification implies the presence of systematic errors and thus invites the same criticism that fell upon the old adaptive methods. Bray and Savin knew this to be a concern and, further, found that the model misspecification may be detectable by agents.

To address this concern, Bullard [2] analyzed learning models in which agents believe the model's parameters to be varying with time.[1] He assumed agents believed the parameters to

*The author would like to thank George Evans and William Branch for helpful comments. All errors are mine.

[1] In fact, Bray and Savin suggested considering models with time-varying parameter beliefs.

be following either a random walk or a stationary AR(1) process, and that the conditional variance of the process was constant. He found that convergence to the REE failed to obtain.[2] We conclude that a necessary condition for convergence is a conditional variance which decreases to zero; and we believe this is a natural assumption to make; indeed, if agents initially use OLS to form their estimates, then the results of Bray and Savin show the conditional variance of the process describing the actual parameters will decrease to zero.

In [6], McGough analyzes Bray and Savin's cobweb model assuming that agents' believe the parameters follow a random walk with decreasing conditional variance. He shows that if the norm of the conditional variance falls slightly faster than $1/t^2$ then convergence to REE obtains. This result leaves open two obvious questions: first, is the obtained bound necessary; and second, can the analysis be extended to cover AR(1) models. As these results appear difficult to address analytically, in this paper we turn to numerical methods. We find that, if agents believe the conditional variance falls like $1/t^e$, convergence to REE appears to obtain for e as low as 1.09 in the random walk model and for e as low as .52 in the AR(1) model. We conclude that the bound obtained by McGough does not appear to be tight, and further, convergence obtains more easily in the AR(1) model.

2 The Model

Bray and Savin's cobweb model is given by $y_t = mx_t + \alpha E_t^* y_t + u_t$, where (x_t, u_t) is *i.i.d.*, $cov(x_t, u_t) = 0$, and, for computational simplicity, all variables are univariate with zero mean. Following Bray and Savin, we assume x_t is observed prior to forming expectations, but that u_t and y_t are not. The asterisk above the expectations operator indicates agents may not be fully rational. In case of full rationality, the unique REE is given by $y_t = \frac{m}{1-\alpha} x_t + u_t$. Agents believe the economy is described by the following model:

$$\begin{aligned} y_t &= \beta_t x_t + \epsilon_t \\ \beta_t - \overline{\beta} &= A(\beta_{t-1} - \overline{\beta}) + \eta_t, \end{aligned}$$

where $var(\eta_t) = 1/t^e$. Denote by b_{t-1} the estimate of β_t at time $t-1$. Then $E_t y_t = b_{t-1} x_t$. This may be used to compute the actual law of motion: $y_t = (m + \alpha b_{t-1}) x_t + u_t$.

Estimation of the agents' beliefs parameters is achieved using the Kalman Filter. Set $\xi_{1t} = \overline{\beta}, \xi_{2t} = \beta_t - \overline{\beta}, z_t = [x_t, x_t]'$. Using the notation

$$x \oplus y = \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix},$$

agents' beliefs may then be written as

$$\begin{aligned} \xi_t &= (I \oplus A)\xi_{t-1} + \begin{bmatrix} 0 \\ \eta_t \end{bmatrix} \\ y_t &= \xi_t' z_t + \epsilon_t. \end{aligned}$$

[2] Margaritis [5] also considered a model with time-varying parameters and constant conditional variance. He obtained a convergence result, but for convergence to a point to obtain, his result required the gain of the adaptive algorithm to tend to zero. As shown by Bullard, this can not occur in our model.

The Kalman filter equations combined with the actual law of motion are written as

$$\begin{aligned} K_t &= (I \oplus A)P_{t-1}\left[\sigma^2 + z_t' P_{t-1} z_t\right]^{-1} \\ \hat{\xi}_t &= (I \oplus A)\hat{\xi}_{t-1} + K_t z_t \left[(\alpha - 1)\hat{\xi}_{t-1}' z_t + [m, 0]\, z_t + u_t\right] \\ P_t &= (I \oplus A)P_{t-1}(I \oplus A)' - \frac{(I \oplus A)P_{t-1} z_t z_t' P_{t-1}(I \oplus A)'}{\sigma^2 + z_t' P_{t-1} z_t} + (0 \oplus var(\eta_t)), \end{aligned}$$

where $b_t = \hat{\xi}_{1t} + \hat{\xi}_{2t}$ is agents' estimate of β_{t+1} at time t. Notice that in case $A = 0$, or $A = 1$ and $var(\eta_t) = 0$, the above algorithm reduces to RLS. The following theorem summarizes what is known.

Theorem 1 *Assume $\alpha < 1$ and $|A| \leq 1$.*

1. *(Bray and Savin) If $A = 0$ and $var(\eta_t)$ is any sequence of positive definite matrices, or if $A = 1$ and $var(\eta_t) = 0$ then $b_t \to \frac{m}{1-\alpha}$ with probability one.*
2. *(Bullard) If $A \neq 0$ and $var(\eta_t) = \Omega$ is positive definite then $b_t \to \frac{m}{1-\alpha}$ with probability zero.*
3. *(McGough) If $A = 1$ and $limsup_t t^2 \|var(\eta_t)\| = 0$ then $b_t \to \frac{m}{1-\alpha}$ with probability one.*

Bray and Savin considered only the case $var(\eta_t) = 0$; however, provided $A = 0$, for any conditional variance the problem reduces to RLS.

3 Convergence Analysis

Numerical analysis of the asymptotic behavior of a simulated stochastic process is difficult because only initial segments of samples are observed, and there are no conditions on initial segments sufficient to guarantee convergence. To avoid this problem, we define what it means for a process to *appear to converge* to a real number using two criteria as developed below.

Let x_t be a stochastic process, $x_t(i)$ a given realization, and x the real number to which we think x_t converges with probability one. Assume we may obtain repeated independent realizations of $\{x_t\}$ of any finite length.

3.1 Criterion #1: Getting Close

A necessary condition for convergence of a sequence to a real number is that the sequence gets close to the number and stays close, at least eventually. The first criterion is based on this necessary condition.

Let $p \in (0, 1)$, $n, M, T, F \in$ N. For each simulation $i \in \{1, \cdots, M\}$ obtain an initial segment of the realization of length n. Set $\varepsilon_i = \frac{p}{n}\sum_{j=1}^{n} |x_j(i) - x|$, and note ε_i is defined as a proportion p of the mean over the initial segment of the distance between $x_t(i)$ and x. Intuitively, we will say that the criterion is passed if before period F, a segment of length T stays in a of x for all M simulations. Formally, we say criterion #1 is met provided $\forall i \in \{1, \cdots, M\}$, $\exists N(i) \leq F - T$ so that $|x_t(i) - x| < \varepsilon_i$ for all $t \in \{N(i), \cdots, N(i) + T\}$.[3]

[3]It may seem natural to also require a realization that has stayed near x for T periods to not escape the

3.2 Criterion #2: Getting Closer

Criterion #1 may be passed even when the sequence of distances to x appears to have stopped decreasing and yet convergence has not obtained. Thus we also determine whether this sequence of distances appears to be decreasing. The intuition for criterion #2 is based on the following observation of Marcet and Sargent, [4]: if $x_t \to x$ in such a way that there exists $\delta > 0$ with $t^\delta(x_t - x) \xrightarrow{D} \hat{F}$ for some nontrivial zero mean distribution $\hat{F}$, then for any $k \in \mathrm{N}$,

$$\frac{E(t^\delta(x_t - x))^2}{E((kt)^\delta(x_{tk} - x))^2} \to 1,$$

so that $E(x_t - x)^2/E(x_{tk} - x)^2 \to k^{2\delta}$. The positive number δ is thought of as the speed of convergence.[4]

Pick $l, t, k \in \mathrm{N}$. Intuitively, if we determine either the sequence has converged to x or that $k^{2\delta} > 1$ then we say that criterion #2 is passed. To test the hypothesis that $k^{2\delta} > 1$ we obtain l realizations of length t and, independently, l realizations of length kt. Set $y_i = (x_t(i) - x)^2$ and $z_i = (x_{kt}(i) - x)^2$, and finally, let $w_i = y_i - z_i$. Notice that $k^{2\delta} > 1$ if and only if $E(w) > 0$. We test the null hypothesis $E(w) \leq 0$ at the α_c-critical level using the asymptotic normality of $\overline{w}$. We say that criterion #2 is passed provided the null hypothesis is rejected.[5]

Definition: The stochastic process x_t *appears to converge* to x provided criteria #1 and #2 are met.

4 Results

Recall $var(\eta_t) = 1/t^e$ so that e measures the rate at which agents believe the conditional variance is decreasing to zero. Letting $\theta = (A, \alpha, m)$, a model may be identified with the parameter pair (θ, e). Because the computational intensity of the criteria prohibits even a coarse analysis of the parameter space, we are only able to make precise statements about a few models (θ, e).

Intuitively, for fixed θ, there should be a boundary value $\rho(\theta)$ so that convergence to REE obtains if and only if $e > \rho$; the intuition being that higher values of e asymptotically yield smaller values of the estimator's gain.[6] We apply this intuition when choosing which values of e to analyze. Specifically, we obtain an estimate $\hat{\rho}$ of ρ as follows: for given criteria

---

ε_i-neighborhood before the time period F is reached. To reduce computational intensity, we do not make such a requirement; this allows us to "pass" a given simulation with out computing all F terms.

[4]We emphasize that the observation of Marcet and Sargent is based on the assumption that convergence obtains; an assumption which we can not make. We use their observation only to provide intuition as to why the ratio $E(x_t - x)^2/E(x_{tk} - x)^2$ might be of interest.

[5]Our criterion # 2 is very similar to the analysis of stopping times as presented by Kenneth Judd, [3]. When determining the point of convergence of a sequence, he advises analyzing the rate of convergence. Our method differs from his method because our problem differs slightly from his; we attempt to show convergence of a stochastic process to a known point, where as he attempts to determine the point of convergence, given a deterministic process known to be converging.

[6]We know of no corresponding formal result; see Appendix for details. Also, of the many simulations run, none violated this intuition.

parameters, we let $\hat{\rho}$ be the lowest value of e analyzed, accurate to the hundredths place, for which convergence appears to obtain 75% of the time; our method is described in detail in the Appendix.[7] We think of $\hat{\rho}$ as an estimate of the conjectured boundary between convergence and non-convergence, but emphasize that our only solid conclusion is $e = \hat{\rho}$ implies apparent convergence 75% of the time.

We choose several benchmark models θ^* and compute $\hat{\rho}(\theta^*)$. We also compute $\hat{\rho}(\theta)$ for models θ obtained by individually varying either α or m from their benchmark values. As these $\hat{\rho}$ do not, in general, appear to differ significantly from $\hat{\rho}(\theta^*)$, we only report $\hat{\rho}(\theta^*)$ and the extreme values of $\hat{\rho}$ obtained by varying α and m. For all experiments, the *i.i.d.* processes x_t and u_t were taken to be standard normal. The criteria parameters used are as follows: $M = 10$; $n = 50$; $p = .01$; $F = 10e7$; $T = 5000$; $l = 100$; $t = 10000$; $k = 4$; and $\alpha_c = 1.96$.

4.1 Random Walk Perceptions

The first benchmark model is $\theta^* = (1, -.5, .5)$.[8] We find $\hat{\rho}(\theta^*) = 1.53$ with extreme estimates $\hat{\rho}(1, -.5, -.5) = 1.49$ and $\hat{\rho}(1, .5, .5) = 1.59$. These findings indicate that the sufficient condition obtained by McGough is not a tight bound; in particular, convergence appears to obtain even when e is significantly less than 2.

To determine whether $\hat{\rho}$ changes for larger variations in θ, we consider the new benchmark $\theta^* = (1, .5, 4)$. We obtain $\hat{\rho}(\theta^*) = 1.13$ with the extreme values $\hat{\rho}(1, .5, 5) = 1.09$ and $\hat{\rho}(1, .9, 4) = 1.52$.[9] These results suggest that ρ may indeed depend on the model's structural parameters. Note too that $\hat{\rho}$ is near one for some values of θ; far lower than the sufficient bound of $e > 2$.

4.2 AR(1) Perceptions

To analyze the AR(1) model we begin with the benchmark $\theta^* = (.9, -.5, .5)$. We find $\hat{\rho}(\theta^*) = .77$ with extreme values given by the benchmark and $\hat{\rho}(.9, -.5, .1) = .76$. This indicates that the $\hat{\rho}$'s associated to AR(1) models may be lower than those associated to the random walk model; perhaps this is because agents have stationary perceptions.

To test dependence of $\hat{\rho}$ on the damping parameter A we consider the new benchmark $\theta^* = (.1, -.5, .5)$. In this case, we find $\hat{\rho}(\theta^*) = .55$ with extreme values $\hat{\rho}(.1, -1.5, .5) = .52$ and $\hat{\rho}(.1, .1, .5) = .57$, indicating a decrease in the damping parameter of the AR(1) process leads to a decrease in $\hat{\rho}$. This is consistent with the theoretical result that $\rho \to 0$ as $A \to 0$.

[7] For any given set of criteria parameters, the boundary between apparent convergence and non-apparent convergence will be grey; this decision rule puts our estimate $\hat{\rho}$ somewhere in the middle of the grey region.

[8] Recall that in the cobweb model, the parameter α represents the ratio of the slopes of supply and demand, and so in the usual case it will be negative.

[9] In this case, increasing the expectations parameter α does seem to significantly impact the estimated value of ρ.

4.3 Example Time-Series

To obtain a feel for the results obtained above we provide three graphs, see Figure 1. Each graph represents a single realization $b_t = \hat{\xi}_{1t} + \hat{\xi}_{2t}$ of length 5000 with $\theta = (1, -.5, .5)$, and initial condition $b_0 = 1$. The REE value is depicted as a horizontal line. In graph (a), $e = 0$ and, consistent with Bullard's result, convergence fails to appear to obtain. In graph (b), $e = 2.5$ and, consistent with McGough's result, convergence appears to obtain. Finally, in graph (c), we take $e = \hat{\rho}(\theta)$, demonstrating the time-series for b_t with e on the conjectured boundary.

[Insert Figure 1]

5 Non-convergence

A question remains: if the process fails to converge to $m/(1-\alpha)$, how does it behave asymptotically? Figure 1, graph (a) suggests that divergence to ∞ is not a necessary result, and in fact, repeated simulations show a process that appears to settle noisily about a fixed mean. Qualitatively similar results are obtained for $0 < e < \rho(e)$. This suggests the following:

Conjecture 2 *If $e < \rho(\theta)$ then $b_t(e)$ is asymptotically stationary and $E(b_t(e)) \to \frac{m}{1-\alpha}$.*

It is not the intent of this paper to statistically analyze this conjecture. However, simple statistical support is not difficult to obtain. Let f_s denote the density of $b_s(0)$. Non-parametric analysis shows that f_t is within a 95% confidence interval of f_s for $s = 10000$ and $t = 20000$, indicating the process b_t may well be asymptotically stationary. Also, the sample mean $\overline{b_t}$ was computed and found not to be statistically different from $m/(1-\alpha)$.

6 Conclusion

The main result of this paper is that in models with PLMs having time-varying parameters, convergence to REE can appear to obtain for rates of decrease in the norm of the conditional variance which are far lower than current analytic results suggest; and further, minimum rates of decrease sufficient for convergence to obtain appear to depend on the specification of the perceived model. Also, notice that if agents believe in an AR(1) process – the process advocated by Bray and Savin as a possible alternative to a constant parameters model – and if A is small then ρ appears to be approximately $1/2$. This is of more than passing importance; we know that if agents use OLS for their estimator then convergence obtains at a rate of $1/\sqrt{t}$. Perhaps, then, if agents detected a misspecification, it would quite natural for them to use $e \sim 1/2$ as their perceived rate of decrease, and consequently guarantee convergence to REE.

Appendix

Let $\Omega(\theta)$ be the set of all $e \geq 0$ such that convergence to REE fails to obtain and let $\rho(\theta) = \sup \Omega(\theta)$. Bullard's results tell us that $0 \in \Omega$ for all θ, making ρ well defined. McGough's result says that if $\alpha < 1$ and $A = 1$ then $\rho(\theta) \leq 2$. We conjecture that if $\alpha < 1$ then the set $\Omega(\theta)$ is a bounded interval. We are unaware of a proof of this conjecture at this time, though no experiments contradicted it. If this conjecture is true then for fixed θ, $e > \rho(\theta)$ would imply convergence and $e < \rho(\theta)$ would imply non-convergence. To estimate $\rho(\theta)$ we first specify the parameters P to be used in the criteria mentioned and then run tests; here a test is taken to be a collection of simulations used to determine whether both criteria are passed. Each value of e is analyzed using four tests. We take our estimate, $\hat{\rho}$, to be the first number e analyzed, accurate to the hundredths place, for which convergence appears to obtain in at least three out of four tests, and so that $e' = e - .01$ implies convergence appears to obtain in fewer than three out of four tests.

References

[1] M. Bray and N. Savin. Rational expectations, learning, and model specification. *Econometrica*, pages 1129 – 1160, 1986.

[2] J. Bullard. Time-varying parameters and non-convergence to rational expectations equilibrium under least squares learning. *Economics Letters*, pages 159 – 166, 1992.

[3] K. Judd. *Numerical Methods in Economics*. MIT Press, 1998.

[4] A. Marcet and T. Sargent. Speed of convergence of least squares learning to rational expectations. In A. Kirman and M. Salmon, editors, *Learning and Rationality in Economics*. Basil Blackwell, Oxford, 1995.

[5] D. Margaritis. A time-varying model of rational learning. *Economics Letters*, pages 309 – 314, 1990.

[6] B. McGough. Statistical learning with time-varying parameters. *Macroeconomic Dynamics*, Forthcomming.

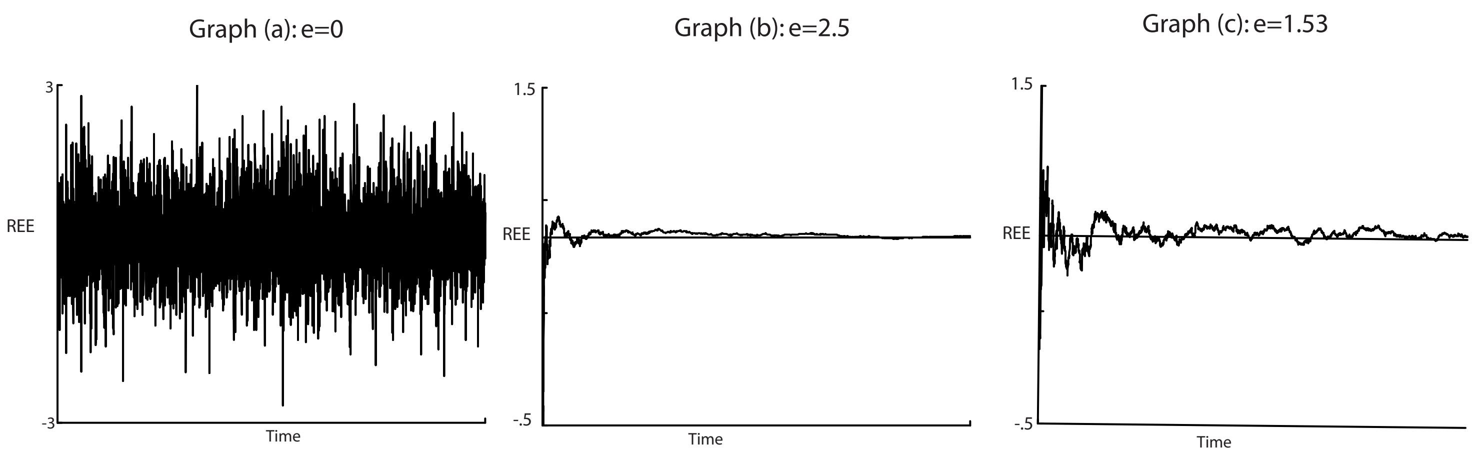
Graph (a): e=0
3
REE
-3
Time
Graph (b): e=2.5
1.5
REE
-.5
Time
Graph (c): e=1.53
1.5
REE
-.5
Time

Figure 1