DEPARTMENT OF ECONOMICS

Working Paper

STABILITY OF FUNCTIONAL RATIONAL
EXPECTATIONS EQUILIBRIA

David L. Kelly
Jamsheed Shorish

Working Paper No. 1999-5

ISSN 1396-2426

UNIVERSITY OF AARHUS • DENMARK

#	INSTITUT FOR ØKONOMI
AFDELING FOR NATIONALØKONOMI - AARHUS UNIVERSITET - BYGNING 350
8000 AARHUS C - ☎ 89 42 11 33 - TELEFAX 86 13 63 34

WORKING PAPER

STABILITY OF FUNCTIONAL RATIONAL EXPECTATIONS EQUILIBRIA

David L. Kelly
Jamsheed Shorish

Working Paper No. 1999-5

DEPARTMENT OF ECONOMICS
SCHOOL OF ECONOMICS AND MANAGEMENT - UNIVERSITY OF AARHUS - BUILDING 350
8000 AARHUS C - DENMARK ☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

Stability of Functional Rational Expectations Equilibria*

David L. Kelly

Jamsheed Shorish

Department of Economics

University of Miami

Box 284126

Coral Gables, FL 33124

dkelly@exchange.sba.miami.edu

and

Department of Economics

University of Aarhus

8000 Aarhus C, Denmark

jshorish@econ.au.dk

First Version: 1996

Current Version: October 27, 1997

Abstract

In this paper we examine a representative agent forecasting prices in a first-order self-referential overlapping generations model. We first consider intermediate stage learning, where agents update the forecasting rule every τ periods. We show that, in theory and simulations, the learning rule does not converge to the rational expectations equilibrium (REE). We next consider two stage learning, where agents learn the functional mapping between the current forecast function and the previous forecast function. We show that in theory and simulations the two stage learning rule converges to the REE.

JEL classification numbers D83, D84, E31. Keywords: functional rational expectations, learning, overlapping generations.

---

* This paper was begun while the authors were at the Graduate School of Industrial Administration at Carnegie Mellon University. We would like to thank Stephen E. Spear for important discussions of this paper. We would also like to thank Ken Koedinger, Duane Seppi, Lael Schooler, Fallaw Sowell, Shyam Sunder and Stan Zin for helpful comments and suggestions, and Beth Shorish for her patience and guidance. Finally, we would like to thank conference and seminar participants at Carnegie Mellon University, the Federal Reserve Board, the Federal Reserve Bank of St. Louis, George Mason University and Queens University. All remaining errors are our own.

The rational expectations hypothesis, in which agents correctly forecast future state variables given current information, is a standard assumption in self-referential economic models. Authors such as Muth (1961) and Lucas (1986) argue that rational expectations is an equilibrium condition. If we believe that agents learn from their environment, then eventually they eliminate all systematic errors—agents truly know how the economy functions, what its laws are, and how it behaves. One might say that under the rational expectations hypothesis, agents 'learn to be rational.' At this point, the economy attains the rational expectations equilibrium (REE). A large body of literature exists demonstrating existence and uniqueness of the REE for a variety of models, and many models take the REE as a natural starting point for describing the state of an economy.

The argument that rational expectations is an equilibrium condition is compelling as long as there exists some kind of procedure which an agent can use to find out everything about the economy that is necessary to forecast without systematic errors. The agent must learn the various relationships between the state variables (and sometimes even what the state variables are) which are dictated by the prevailing model, even if such a model is mathematically complex, before any hope can be found of reaching the REE. Whether or not such a procedure exists has been studied extensively in the literature. Following Spear (1989) we can divide the literature into four categories, Bayesian learning and three versions of adaptive learning from the state variables.

Several papers, including Blume and Easley (1981), Cyert and Degroot (1974), Easley and Kiefer (1989), El-Gamal and Sundaram (1989), Feldman (1987), Nyarko (1991) and Townsend (1983) focus on Bayesian learning. In these models, agents begin with some initial prior distribution assumption over the future state variables (sometimes with heterogeneous priors across agents). By observing the forecasts of others and/or market observable (e.g. prices) agents infer information about the true distribution of the uncertain variables, and use Bayesian updating to modify the parameters of the prior distribution. When every agent infers all information, the economy attains the REE. Agents are shown to learn the REE in linear models (see e.g. Townsend, 1983).

Bayesian learning is especially important because it is not an ad hoc assumption, but an actual model of learning. Bayesian learning, however, can be difficult to apply if there is more than one uncertain variable, or under prior distributions which are not normally or binomially distributed, since the updated distribution does not in general have the same functional form as the prior.

Adaptive learning methods, based on realizations of the state variables, typically utilize some form of a "forecast function" which, given a vector of current state variables, generates a forecast of the future state variables. Once the future becomes known, the forecast function may be changed to reflect the new data. Following Spear (1989) we call the first category of adaptive learning "incremental learning," in which agents update their forecast function after every observation of the state variables. Agents learn to be rational if the sequence of forecast functions over time converges to the REE, i.e. the forecast function reflects the true law of motion that dictates the future state variables. Outside the REE, i.e. in the temporary equilibrium (see Grandmont 1977), the forecast functions are not rational. Incremental learning is advantageous since agents use all available information to update each forecast function. Bray and Savin (1986), Marcet and Sargent (1989a, 1989b) and others show that least squares learning, applied incrementally, can converge to the REE in linear models. However, Bullard and Duffy (1993) show that this result does not generally hold when agents make forecasts that are more than 3 steps ahead.

If each update of the forecast function is costly, however, then incremental learning may be prohibitively expensive for agents to employ. In addition, as pointed out in Spear (1989), agents may require several bad forecasts to realize that the forecast function is systematically wrong. In this case, agents may not update the forecast function in a given period, believing the economy is at the REE when it is not. Hence we might expect that agents accumulate observations before updating.

We let "intermediate stage learning" denote an alternative to incremental learning, where agents accumulate observations on the state variables for τ periods before updating the forecast function. Since agents do not update the forecast function each period, this procedure effectively fixes the temporary equilibrium function (TEF) in between updates. Hence, given enough realizations of the state variables, the TEF could be learned, which would give the agent the law of motion of the state variables up to that point. Agents would then use the TEF as their new forecast function. Unfortunately, this changes the TEF and agents must again accumulate observations to learn the new law of motion. If this sequence of forecast functions converges to the REE, agents have learned to be rational. Brock (1972), DeCanio (1979) and others have shown that intermediate stage learning converges to the REE in linear self-referential models.

Since intermediate stage learning assumes that the temporary equilibrium function is learned, we find that an agent may have an especially difficult task in models with non-linear laws of motion. Another consideration is that an agent needs to know the REE in order to calculate whether or not further updating is optimal. Therefore, τ must be set in an *ad hoc* manner. On the other hand, an agent may realize that submitting a new forecast function changes the temporary equilibrium of the model, which is continually used as the *new* forecast function after updating. In this case, agents may try to learn the *functional* mapping that takes current forecast functions into future forecast functions.

Agents that attempt to learn this functional mapping use "two stage learning." In the first stage of this process, agents accumulate pairs of current and future forecast functions. In the second stage, agents try to learn the functional mapping between these forecast functions. Spear (1989) shows that learning procedures which converge to the REE exist using two stage learning with full information. Board (1992), however, shows that for a wide class of models, agents cannot learn to be rational in a finite time (under the "probably approximately correct" learning axiom). Furthermore, finding an algorithm that systematically updates in the function space is difficult—agents must submit many errant forecasts to obtain each forecast function pair. Hence agents on their way to rationality may pay a large price in utility from these errant forecasts. Finally, even when the functional mapping is learned, the agents must still compute the mapping's functional fixed point, which is the REE.

In this paper, we introduce a general learning procedure which can be used in an intermediate stage or two stage manner. We show that for a class of non-linear general equilibrium models, agents do not learn to be rational when using intermediate stage learning. This result appears to depend crucially on the two step ahead forecasting rule used—the current values of the state variables depend upon the agent's forecast of the *future* state variables. However, for the same class of non-linear models we show that two stage learning converges to the REE.

We demonstrate the theoretical results outlined above by simulating linear and non-linear economies with intermediate stage and two stage learning. We introduce a method of learning which can learn functional mappings as well as continuous functions. This method utilizes functions which are drawn from a class of 'universal approximators': Fourier analysis, Kolmogorov functions, and Radial Basis Functions are all examples from this class. We select neural networks from this class, and generate algorithms which are used to demonstrate intermediate and two stage learning. We show that for intermediate stage learning a neural network successfully learns the non-linear temporary equilibrium, but the sequence of forecast functions using the TEF diverges away from the REE. In two-stage learning, neural networks successfully learn the functional mapping between current and future forecast functions. Furthermore, the fixed point of the functional mapping (the REE) is shown to be easily computable using established methods.

The chapter is divided into the following sections. Section 1 introduces a first-order self-referential economy derived from a stochastic overlapping generations model, and the REE of the economy is derived. In Section 2 the theoretical results for intermediate stage and two stage learning are presented. Section 3 gives the results for simulations of the models outlined in Sections 1 and 2, using the neural network learning algorithm. Section 4 presents concluding remarks. Proofs of Propositions, Theorems, etc. are in the Appendix.

1 The Model

1.1 The Overlapping Generations Framework

We consider an overlapping generations (OLG) model with uncertainty. Let time be indexed as t = 0, 1, … There is a continuum of agents of two types (so that the representative agent formulation applies to each type), and all agents live for two periods.

There exists a single perishable commodity which is consumed. Uncertainty enters the model through w_t, the agent's endowment when young. We let the young agent's endowment be distributed according to $w_t \sim G$ i.i.d., with G a continuous distribution function over a compact set $\Omega \subset \Re$. The young agent observes the current endowment before any decision is made. Therefore, the only uncertainty is over the realization of the endowment by the next young generation, which affects the price when the young agent becomes old. Old agents receive a non-stochastic endowment w_2.

We assume that there exists a storable asset, fiat money, which the agents use to transfer wealth between youth and old age. There is a fixed supply of money $\overline{m}$ endowed to the old agents born at time zero.

Agents have preferences over consumption in the young and old periods. We define a concave, twice differentiable utility function $U(c_t, c_{t+1})$. Agents may trade consumption goods for money on a spot market at price p_t. We assume a compact price space $S \subset \Re^+$ and we let money be the numeraire good. The young agent maximizes expected utility subject to a budget constraint:

$$\max_{c_t, c_{t+1}, m_t} E_t U(c_t, c_{t+1}) \qquad 1.1.1$$

subject to

$$c_t = w_t - \frac{m_t}{p_t} \qquad 1.1.2$$

$$c_{t+1} = w_2 + \frac{m_t}{p_{t+1}^e} \qquad 1.1.3$$

Here p_{t+1}^e is the point forecast of the price expected to prevail in period t + 1. Since we will be dealing with adaptive learning methods with forecast functions in what follows, we note here that we are imposing a behavioral condition on the agent which is different than, for example, the Bayesian procedure in which the agent assumes a distribution function over p_{t+1}^{e}. Here the agent forecasts the future by submitting a guess of the future price to a Walrasian auctioneer. Since p_{t+1}^{e} is the expected price for period *t+1*, it will naturally depend upon the endowment realization in period *t+1*.

Substituting the constraints into the utility function yields the following maximization problem:

$$\max_{m_t} \int U\left(w_t - \frac{m_t}{p_t}, w_2 + \frac{m_t}{p_{t+1}^{e}} \right) dG(w_{t+1}). \qquad 1.1.4$$

This problem has a first order condition given by

$$\int \left(-\frac{1}{p_t} U_1 + \frac{1}{p_{t+1}^{e}} U_2 \right) dG(w_{t+1}) = 0, \qquad 1.1.5$$

where $U_i, i = 1,2$ is the derivative of *U* with respect to the i[th] argument.

Since in equilibrium money supply and money demand are equal, we can write the first order condition 1.1.5 as

$$\int z\left(p_t, w_t, p_{t+1}^{e}\right) dG(w_{t+1}) = 0 \qquad 1.1.6$$

(for brevity we have suppressed *z*'s dependence on the total money stock $\overline{m}$). Equation 1.1.6 is the law of motion for the economy. We note that there exist many other models that can be reduced to equation 1.1.6 with p_t typically representing prices, or supply of capital or labor.[1] The above model also illustrates a difficult problem for the agent—in order to learn the REE, the agent must *learn* equation 1.1.6, which may be a complex, non-linear, stochastic process in which forecasts (p_{t+1}^{e}) affect current values of the state variables (here, p_t). To make matters even worse, what the agent learns may qualitatively affect the equilibrium of the OLG model. For example, Grandmont (1985) shows that for different adaptive learning rules, endogenous business cycles of varying periodicity can occur in OLG models.

1.2 Forecasting Rules

We assume the agent uses a first–order forecasting rule, which mimics the first-order dependence of p_{t+1}^{e} on p_t in the temporary equilibrium function 1.1.6. This may appear to

[1] See e.g. Azariadis (1981), Grandmont (1985) and Marcet and Sargent (1989a).

be a restrictive assumption on the class of forecast functions, but the results presented go through even if there are other variables (e.g. higher-order lags, or sunspot variables) within the forecast function, due to the nature of 1.1.6. We thus restrict ourselves to first-order dependence for expositional simplicity.[2] Next, it is important to note that when the agent forecasts p^e_{t+1}, the actual current price p_t has yet to be determined (in this simple OLG model, the price is determined by a Walrasian auctioneer). In fact, the agent takes p_t as given when maximizing utility, and must forecast p_{t+1} using the last available data point, p_{t-1}. However, a first-order forecasting rule allows the agent to update the forecast function once the current price p_t is revealed—the agent still has no information about p_{t+1}, but now has the data (p_{t-1}, p_t).

An example may help clarify matters. Agent Sally supposes that the current price p_t (which is not yet observed) depends only upon the previous known price p_{t-1}, and thus has a forecast function which gives the expected *current* price as a function of p_{t-1}:

$$p^e_t = g(p_{t-1}).$$

Now Sally is called upon to form an expectation about the *future* price p_{t+1}. Using her first-order forecast function, her expected future price is given by

$$p^e_{t+1} = g\left(p^e_t\right) = g\left(g\left(p_{t-1}\right)\right).$$

Sally submits this p^e_{t+1} (which depends only upon p_{t-1}) to the auctioneer, who then calculates the current price p_t. Now Sally is free to update the function *g* since data is available on both p_t and p_{t-1}. For example, Sally could modify *g* in some way by minimizing the error $\left(p_t - p^e_t\right)^2 = \left(p_t - g\left(p_{t-1}\right)\right)^2$. Thus, the forecast function *g* can incorporate new information, even though nothing is known about p_{t+1}.

On to formalities. We want to keep the analysis grounded in what an economic agent would be expected to achieve in reality. Thus, we suppose that our first-order forecast function is parameterized by a finite vector of real numbers, as any forecast function used in a real economy must be. Again, the convergence results to be shown do not depend upon such a parameterization; indeed, the proofs simply assume a continuous first-order forecast

[2] A single price lag in the information set can also be interpreted as a "minimum state variable" condition (McCallum, 1989).

function. But to keep the exposition as clear as possible (and minimize further switching around of notation!) we will assume a parameterized forecast function in what follows.

We let B be a compact subset of $\Re^n$ and let β_t be an element of B. We then define a continuous function $f: S \times \Omega \times B \rightarrow S$ such that

$$p_t^e = f(p_{t-1}, w_t, \beta_t). \qquad 1.2.1$$

We note that f is misspecified since the agent is not aware that forecasts influence the state variables, similar to Marcet and Sargent (1989b). This also abstracts away from the question of strategic forecasting.[3]

Substituting the forecast function into the first order condition gives

$$\int\Bigg[-\frac{1}{p_t}U_1\left(w_t - \frac{\overline{m}}{p_t}, w_2 + \frac{\overline{m}}{f(f(p_{t-1}, w_t, \beta_t), w_{t+1}, \beta_t)}\right) + \frac{1}{f(f(p_{t-1}, w_t, \beta_t), w_{t+1}, \beta_t)}U_2\left(w_t - \frac{\overline{m}}{p_t}, w_2 + \frac{\overline{m}}{f(f(p_{t-1}, w_t, \beta_t), w_{t+1}, \beta\ )}\right)\Bigg]dG(w_{t+1}) = 0. \qquad 1.2.2$$

As before we condense the above equation to

$$\int z(p_t, w_t; f(f(p_{t-1}, w_t, \beta_t), w_{t+1}, \beta_t))dG(w_{t+1}) = 0. \qquad 1.2.3$$

Equation 1.2.3 is the stochastic temporary equilibrium function (TEF) of the economy. Note that the function z is functionally dependent on f. For example, if f depended on a second lag of prices, the TEF would be second order.

We next summarize the restrictions made so far:

Assumption A1: *Prices and endowments are defined over the compact sets S and Ω, and the function z is C^1.*

Typically, the endowment space is defined such that a positive monetary stationary state exists. For this to occur, the endowment when old must be small enough to induce the young to trade goods for money. Note that we impost no linearity restriction on the model. Another important consideration is that the agent does not know anything about functional form of the REE. Previous results on learning REE such as Bayesian learning (Feldman, 1987), least squares learning (Marcet and Sargent, 1989b) and genetic algorithms (Arifovic, 1989)

[3] In many (finite agent) models agents may wish to submit strategic forecasts to manipulate current prices if they are aware that forecasts affect prices.

assume that the functional form of the REE is known up to a few parameters. Here the agent must learn the correct function out of the entire space of continuous functions.

1.3 Rational Expectations Equilibrium

A rational expectations equilibrium is a forecast function $p_t^e = f(p_{t-1}, w_t, \beta^*)$ that solves the law of motion, which means that the rational forecast is correct 'on average.' Equilibrium occurs when the implicit function between p_t and p_{t-1} given by equation 1.2.3 is the same as the forecast function the agent is using.

Definition 1: *A Rational Expectations Forecasting Rule is a forecasting rule* $p_t^e = f(p_{t-1}, w_t, \beta^*)$ *such that locally,*

$$\int z\left(f\left(p_{t-1}, w_t, \beta^*\right), w_t; f\left(f\left(p_{t-1}, w_t, \beta^*\right), w_{t+1}, \beta^*\right)\right) dG(w_{t+1}) = 0 . \qquad 1.3.1$$

Existence of the REE follows from the Implicit Function Theorem.

2 Learning and Convergence to the REE

2.1 Intermediate Stage Learning

One way to model learning is through *intermediate stage learning.*

Definition 2: *Intermediate stage learning allows the forecast function to be updated only every* $\tau > 1$ *periods:*

$$\lfloor \beta_{m\tau}, \ldots, \beta_{(m+1)\tau-1} \rfloor \equiv \beta_m, m = 0,1,\ldots \qquad 2.1.1$$

If agents use intermediate stage learning then they are not necessarily trying to make the best forecast every period—incremental learning might then be used. But if it is costly to recalculate the forecast function every period, an agent may opt to defer updating for several periods at a time (see e.g. Evans and Ramsey 1992).

Given p_{t-1}, agents forecast using their initial forecast rule to get p_{t+1}^e, which generates p_t from equation 1.2.3. Agents then have one data point $[p_{t-1}, p_t]$. Agents repeat the process until τ such data points are accumulated. Formally we have:

Definition 3: *A data set for updating the forecast function* $p_t^e = f\left(p_{t-1}, w_t, \beta_\tau\right)$ *is a set* $\{[p_{t-\tau}, \ldots, p_t], [w_{t-\tau}, \ldots, w_t]\}$ *such that*

$$\int z\left(p_k, w_k; f\left(f\left(p_{k-1}, w_k, \beta_\tau\right), w_{k+1}, \beta_\tau\right)\right) dG(w_{k+1}) = 0 \quad \forall k = t - \tau \ldots t . \qquad 2.1.2$$

After τ periods the agent updates the forecast function. The mapping between the old and new forecast functions is defined according to the following equation:

$$\beta_{\tau+1} = \arg\min_{\beta \in B} \left\| p_t - f\left(p_{t-1}, w_t, \beta\right) \right\|. \qquad 2.1.3$$

Here the norm is over the vector of τ data points.

We must now venture back into reality and discuss how this type of forecasting may be adversely affected by 1) bad draws from the environment, and 2) poor performance from the updating procedure itself. As is the case with stochastic linear learning models (e.g. Marcet and Sargent 1989a and 1989b), any convergence to the REE will be a local result. Hence there is the possibility that the $\left[p_{t-1}, p_t\right]$ data set is so far from the mean that the updated forecast function moves outside the domain of attraction (if it exists) of the REE function. We thus restrict our attention to local convergence results in what follows. Unlike linear models, however, there is also the possibility that the forecast function does not adequately represent the temporary equilibrium when the minimization of 2.1.3 is performed. Again, the updated forecast function may find itself outside the domain of attraction (if it exists) of the REE function.

Naturally, if the REE is locally unstable under learning, then these problems are irrelevant—the learning routine will not converge in any case. But for those cases in which the REE would be stable if learning were perfect, it is necessary to restrict the class of REE to those for which small perturbations in the temporary equilibrium function do not change the dynamics around the REE. This may exclude convergence to many other types of equilibria. For example, a model for which perturbations in the TEF generate bifurcations in the set of REE, e.g. a stable, unique REE may under *perturbed* learning become an unstable member of a 3-element equilibrium set.

However, we wish to stress that this says nothing about whether a learning system will or will not converge to an REE which is restricted in the above sense. An REE which does not change its dynamics under small perturbations of the TEF might be stable or unstable under intermediate or (as defined later) two stage learning. The restriction, outlined formally in Assumption A2 below, is simply used focus attention on those REEs for which an agent using a realistic forecasting routine might be expected to learn.

Assumption A2: *Suppose Definition 2 holds. Define a forecast function approximation error as function* η *such that at any time*

$$\eta = \left\| p_t - f\left(p_{t-1}, w_t, \beta_\tau\right) \right\|.$$

Suppose that the economy is at a point where the agent updates. We rewrite the law of motion as:

$$\int z\left(f\left(p_{t-1}, w_t, \beta_{\tau+1}\right) - \eta, w_t; f\left(f\left(p_{t-1}, w_t, \beta_\tau\right), w_{t+1}, \beta_\tau\right)\right) dG\left(w_{t+1}\right) = 0 . \qquad 2.1.4$$

Then the dynamic stability of the system 2.1.4 *is identical to the system where* $\eta = 0$ *:*

$$\int z\left(f\left(p_{t-1}, w_t, \beta_{\tau+1}\right), w_t; f\left(f\left(p_{t-1}, w_t, \beta_\tau\right), w_{t+1}, \beta_\tau\right)\right) dG\left(w_{t+1}\right) = 0 . \qquad 2.1.5$$

Since we are working with functional REE, the functional notation will be particularly useful in what follows. To rewrite the problem to a functional one, we define an operator $T : C^1_{S\times\Omega\times B} \to C^1_{S\times\Omega\times B}$ implicitly as

$$\int z\left(T(f), w_t; f\left(f\left(p_{t-1}, w_t, \beta_\tau\right), w_{t+1}, \beta_\tau\right)\right) dG\left(w_{t+1}\right) = 0 . \qquad 2.1.6$$

Given the collected data, by A2 we can restrict our attention to agents who estimate the TEF accurately, i.e.

$$f\left(p_{t-1}, \beta_{\tau+1}\right) = T \circ f\left(p_{t-1}, \beta_\tau\right). \qquad 2.1.7$$

T maps current forecast functions into future forecast functions. From equation 2.1.6 *T* is at least piecewise continuous. Since the forecast function is first-order, we have seen that agents must iterate their old forecast function to generate the new forecasting function:

$$p^e_{t+1} = f_{\tau+1} = T \circ f_\tau \circ T \circ f_\tau , \qquad 2.1.8$$

where we define $f_\tau \equiv f\left(p_{t-1}, w_t, \beta_\tau\right)$. (Since we are working in the function space, we will also drop the arguments of *f* for brevity.) Similarly, we let $f^* \equiv f\left(p_{t-1}, w_t, \beta^*\right)$, so that f^* is our shorthand for the rational expectations equilibrium. We are now in a position to present the main results. The following theorem shows that the sequence of forecast functions defined above diverges away from the rational expectations equilibrium .

THEOREM 1: *Suppose that A1 and A2 are satisfied. Let* $f_0 : S\times\Omega\times B \to S\times\Omega\times B$ *be such that* $f_0 \in C^1_{S\times\Omega\times B}$ *and*

$$f_0 \in N_\varepsilon\left(f^*\right) \equiv \left\{ f : \sup_{p\in S, w\in\Omega} \left| f - f^* \right| < \varepsilon \right\} \qquad 2.1.9$$

for a given ε. Then if f_τ *evolves according to equation 2.1.8, the sequence of forecast functions does not converge (in the sup norm) to* f^* *for any* $f_0 \in N_\varepsilon$*. That is, the rational expectations equilibrium is unstable.*

We note that there are no restrictions on the law of motion other than continuity and invertibility. Hence Theorem 1 applies to a wide range of economic models. The reason behind the divergence is that the two step mapping in the function space (i.e., the operator which results when *T* and *f* are iterated twice) is no longer a contraction. Assuming that the functional mapping is a contraction is an implied assumption in the linear learning literature, as well as in White (1989). The nonconvergence result of Theorem 1 is not unlike that of Bullard and Duffy (1993) where forecasts which are several steps ahead cause some roots to be outside the unit circle and hence cause linear learning rules to become unstable.

We also note that the above result is more a negative result for the learning procedure than for the rational expectations hypothesis. Although intermediate stage learning is an intuitive way for agents to update, with each update the agent forecasts less accurately. Therefore, the utility cost of using intermediate stage learning increases to the point where the agent ought to try something else.

We next test how robust Theorem 1 is to variations of intermediate stage learning. Specifically, we suppose that the agent chooses a new forecast function $f_{\tau+1}$, as a C^1 function of the temporary equilibrium and the most recent forecast function:

$$f_{\tau+1} = \Psi(T \circ f_\tau ; f_\tau). \qquad 2.1.10$$

For example, the agent may choose $f_{\tau+1}$ as a convex combination of the TEF and the current forecast function:

$$f_{\tau+1} = \gamma T(f_\tau) + (1-\gamma) f_\tau. \qquad 2.1.11$$

The following theorem generalizes Theorem 1 to the variant 2.1.10 of intermediate stage learning.

THEOREM 2: *Suppose A1 and A2 are satisfied. Let* f_0 *be as in Theorem 1. Suppose the following consistency condition is satisfied:*

$$\Psi(f^*; f^*) = f^*. \qquad 2.1.12$$

Then if f_τ *evolves according to equation 2.1.10, the sequence of forecast functions does not converge (in the sup norm) to* f^* *for any* $f_0 \in N_\varepsilon$ *and for ε sufficiently small. That is, the rational expectations forecast function is unstable.*

The consistency condition in Theorem 2 states that the agent does not update away from the rational expectations forecast function once the current forecast function is rational (i.e. the agent knows when the REE is achieved).[4] Such a condition is already present in the form of intermediate stage learning given by 2.1.8.

2.2 Two Stage Learning

Given the law of motion 1.1.6, when agents use intermediate stage learning, they cannot learn the rational expectations equilibrium. In this section we describe another learning procedure, called two stage learning, which enables an agent to learn the REE.

Definition 5: *A* two stage learning process *is a learning algorithm which, given a data set consisting of* $(f_\tau, f_{\tau+1})$ *pairs, learns the operator* T *defined implicitly by equation 2.1.6.*

With two stage learning, an agent attempts to learn the *functional* mapping between current and future forecast functions in the C^1 function space.

The two stage learning procedure is composed of the following stages:

Stage One

- An agent discovers that the current forecast function influences the TEF, possibly as a result of trying some intermediate stage learning process.
- After submitting a forecast f_τ, the agent uses intermediate stage learning to learn the temporary equilibrium function $f_{\tau+1}$. The agent thus collects pairs of forecast functions $(f_\tau, f_{\tau+1})$.

Stage Two

- The agent learns the operator *T*, which maps the initial forecast function f_τ into the new forecast function $f_{\tau+1}$.
- The agent then computes the fixed point of *T*, which is the rational expectations equilibrium f^*.

We now develop an algorithm to demonstrate that two stage learning is feasible, using the above stages as guidelines. The following Lemma introduce the existence of one subset of the class of so-called 'universal approximators,' which are groups of functions that are able to approximate to an arbitrary degree of accuracy any measurable function.

[4] See Grandmont (1985) for the consistency condition in the price space.

LEMMA 1 (Hornik, Stinchcombe and White, 1989 and 1990): *For every function* $f \in C_S^1$, $f: S \to S$ *there exists a function* $k \in C_{S \times R^n}^1$, $k: S \times \Re^n \to S$, *a vector* $\beta^f \in \Re^n$ *and an* $\bar{\varepsilon} > 0$ *such that* $\forall \varepsilon > 0$,

$$\left\| k\left(p, \beta^f\right) - f(p) \right\| < \varepsilon, \qquad 2.2.1$$

$$\left\| k_p\left(p, \beta^f\right) - f'(p) \right\| < \bar{\varepsilon}, \qquad 2.2.2$$

where k_p *denotes the derivative of the function* k *with respect to the price* p.

We now recast the two stage learning process in terms of the vectors β which define the forecast functions. Suppose that the agent begins with an initial forecast k_τ parameterized by a vector β_τ. The agent then samples from the price space to learn an approximation of $T \circ k_\tau$, defined as $k_{\tau+1}$ which is parameterized by $\beta_{\tau+1}$. Now the agent has two functions, represented by two parameter vectors β_τ and $\beta_{\tau+1}$. This process is repeated until the agent acquires a series of data vectors $[\beta_\tau, \beta_{\tau+1}]$. This completes the first stage of learning.

In the second stage, the agent attempts to learn the operator *T* between the forecast functions k_τ and $k_{\tau+1}$. We show in the following theorems that learning a vector mapping between β_t and β_{t+1} is sufficient for learning *T*, which is a useful and powerful result as it means that functional mappings can in general be approximated (theoretically to an arbitrary degree of accuracy) by vector-valued mappings. We first generalize Lemma 1 to functional operators:

PROPOSITION 1: *Let* $T: C_{S\times\Omega\times B}^1 \to C_{S\times\Omega\times B}^1$ *be as in Section 2.2. Let A1 and A2 hold. Then there exists an operator* $\hat{T}: C_{S\times\Omega\times B}^1 \times \Re^n \to C_{S\times\Omega\times B}^1$ *such that* $\forall f \in C_{S\times\Omega\times B}^1$, *there exists a vector* $\beta^T \in \Re^n$ *such that for a given* $\varepsilon > 0$,

$$\left\| T(f) - \hat{T}\left(f, \beta^T\right) \right\| < \varepsilon. \qquad 2.2.3$$

Furthermore,

$$\left| \, \left\| \hat{T}'(f) \right\| - \left\| T'(f) \right\| \, \right| < 2\varepsilon. \qquad 2.2.4$$

Proposition 1 shows that there exists an approximation of the functional operator *T* which is parameterized by a vector—in order to show that an algorithm exists which can converge to this approximation, we require the next set of theorems.

THEOREM 3: *Suppose that the REE,* $f^* = T(f^*)$ *satisfies* $\|T'(f^*)\| \neq 0$, *i.e. the REE is a regular value of the operator* T. *Then there exist a vector* $\beta^* \in \Re^n$ *and a function* $k(p,\beta^*) \equiv k^*$ *such that*

$$\hat{T}(k^*) = k^*,$$

$$\lim_{\varepsilon \to 0} \|k^* - f^*\| = 0,$$

$$\|k^* - f^*\| = \varepsilon^*,$$

where

$$\|\hat{T}'(f^*)\| < 1 \Rightarrow \varepsilon^* < \varepsilon\left(\frac{1}{\left(1 - \|\hat{T}'(f^*)\|\right)}\right),$$

$$\|\hat{T}'(f^*)\| > 1 \Rightarrow \varepsilon^* < \varepsilon\left(\frac{\|\hat{T}'(f^*)\|}{\|T'(f^*)\|\left(\|\hat{T}'(f^*)\| - 1\right)}\right).$$

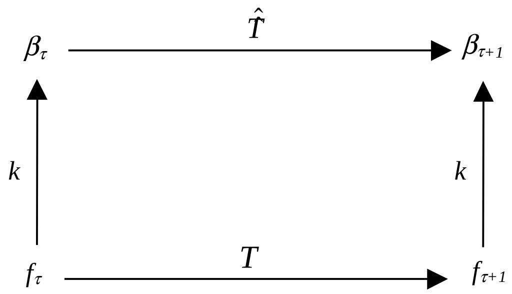
COROLLARY 1: *Suppose* $\|\hat{T}'(f^*)\| \geq 1 + \varepsilon + (1+\varepsilon^2)^{\frac{1}{2}}$. *Then* $\varepsilon^* < \varepsilon$.

Theorem 3 and Corollary 1 say several things—they are the heart of the procedure for approximating the functional mapping *T*. Theorem 3 states first of all that an approximation of the REE f^* exists, which is a fixed point of the vector-parameterized approximation $\hat{T}$ of the operator *T*. This amounts to saying that finding the fixed point of $\hat{T}$ is akin to finding f^*. In addition, the theorem and the following corollary place boundaries on the approximation error between k^* and f^*, which is after all what we are ultimately concerned with. If we presume that the approximation error between T and $\hat{T}$ is 'small enough,' then Corollary 1 ensures that the error between k^* and f^* is even smaller.

Theorem 3 says nothing, however, about actually finding the function k^*. We seem to have traded one functional fixed point search for another. Fortunately, the fact that k^* is parameterized by a vector β^* allows us to define in Theorem 4 a vector mapping with β^* as the fixed point. We can then apply conventional vector fixed point search algorithms to recover β^* and hence k^*.

THEOREM 4: *Let the assumptions of Theorem 4 hold. Then there exists a vector mapping* $\hat{\hat{T}}: \Re^n \to \Re^n$ *such that* $\hat{\hat{T}}(\beta^*) = \beta^*$, *where* β^* *satisfies* $k(p,\beta^*) = k^*$ *and* $k: S \times \Re^n \to S$.

From Lemma 1, the mapping $\hat{\hat{T}}$ can be approximated using the second stage of two stage learning. Then we know that the vector fixed point of the approximation $\hat{\hat{T}}$ defines the mapping k^* which is in an ε-neighborhood of the rational expectations equilibrium f^*. As noted earlier, since $\hat{\hat{T}}$ maps vectors into vectors, the fixed point can be found by an algorithm such as the Gauss-Newton method.[5] The following diagram summarizes the relationship between the forecast functions, the functional operator *T*, and their vector mapping equivalents:



To summarize, we have shown that in theory agents can learn the REE with two stage learning. The key to this procedure is that agents are able to code forecast functions into vectors. Then the problem reduces to one of finding a fixed point in a vector space instead of a function space. Theorem 4 shows that the fixed point of the vector mapping, when decoded, is the approximation of the rational expectations equilibrium.

3 Simulations

We next test the theory presented in Section 2. As defined in Lemma 1, we have need of a vector-valued function approximator belonging to the class of universal approximators. We select neural networks as our universal approximator satisfying Lemma 1 (indeed, Lemma 1 was presented in the context of showing that neural networks are universal approximators—see Hornik *et al*. 1989 and 1990). For the simulations we chose several different functions for the law of motion, which were deterministic to reduce computation time. We note here that the neural network tends to learn linear functions, non-monotonic polynomial functions, and transcendental functions faster than monotonic concave or convex

---

[5] The two stage learning algorithm can also be used to approximate other functional mappings, such as Bellman's equation in dynamic programming problems.

functions. We thus selected laws of motion that the neural network learns both quickly and slowly.

3.1 Simulated Intermediate Stage Learning Network

With intermediate stage learning, agents update their forecast function every τ periods, learning the temporary equilibrium of the economy. However, the results of Section 2.1 predict that the sequence of forecast functions derived from intermediate stage learning does not converge to the rational expectations forecast function, but instead diverges. To show this numerically, we first set up a neural network to learn the rational expectations forecast function f^*. The resulting approximation was then very close (in terms of the sum of squared errors), but not exactly equal, to the true law of motion. Hence this approximation makes a good starting point for the sequence of temporary forecast functions (the f_0) since the function is in a small neighborhood of the rational expectations equilibrium.

Given f_0, we then calculated a training set by dividing the domain equally into 50 points. We then calculated the output $T \circ f_0$ for each of the 50 points. (We designed the law of motion so that we could always solve the first order condition for the *T* operator.) This became the training set for f_1. Using the same input, we then calculated an output set using $T \circ f_1$. This became the training set for f_2 and so on.

For our law of motion under intermediate stage learning we selected a quadratic function

$$p_t = \left(p_{t+1}^e\right)^2, \tag{3.1.1}$$

which generates the functional REE

$$p_{t+1} = \left(p_t\right)^{\frac{1}{2}}. \tag{3.1.2}$$

This law of motion can also be derived from a non-stochastic OLG model.[6] The price domain was the interval $[0,5]$.

For the law of motion defined by equation 3.1.1, we set up a network with n = 4. We found in this case that the sigma function which produced the fastest results was the hyperbolic tangent function, tanh(*x*). We used the Nguyen and Widrow (1990) initial condition algorithm for the f_0 approximation. For the rest of the approximations we let the

---

[6] See e.g. Cass *et al*. (1979).

final f_0 weights be the initial weights for the next run. This increased the rate of convergence. We found that although the initial forecast function was very close to the REE, the sequence of forecast functions generated from the TEFs diverged towards infinity. This is in keeping with the theoretical results presented in Section 2.1. Figures 1-6 show this divergence, with the REE given by equation 3.1.2.

3.2 Simulated Two Stage Learning Network

To simulate two stage learning, we first used the neural network to convert input and output functions to vectors. A neural network then learned the mapping between these vectors. The fixed point of this mapping was then the approximation of the REE, once this vector was converted back into a neural network. We selected the quadratic function (see equation 3.1.1) and also selected a linear law of motion given by

$$f(p_t) = \frac{1}{3} p_t - \frac{4}{3}. \qquad 3.2.1$$

We again used the interval $[0,5]$ as the domain.

For each law of motion we first trained the neural network on the rational expectations equilibrium. We then took the final weights and perturbed them 50 times for the linear law of motion and 100 times for the quadratic law of motion. These were taken as the input forecast functions.[7] Next we took the perturbed weights, created neural networks from these weights, and then ran the input data through $T(f(p_{t-1}, \beta_\tau))$ to get output data points. On these input-output vectors several neural networks were trained, which generated a vector approximation for each $T(f_i)$, where i indexes the 50 (100 in the quadratic case) perturbations. This gave us an input and output vector for each perturbation of the equilibrium function. We then trained a neural network on the 50 (100) input and output vector pairs, as an approximation of the functional operator *T*. Finally, to test if the resulting approximation could correctly identify the rational expectations equilibrium, we ran the approximation of the rational expectations equilibrium (k^*) through the neural network approximation of the operator *T*. If the REE was to be identified, the output of the network would be the same as the input. In addition, we used numerical techniques to approximate the fixed point of the vector mapping directly.

[7] These are local perturbations around the fixed point of the *T* operator in the function space.

For the linear case we set n = 2. With two stage learning, a large value of n not only makes the first stage of learning slower since there are more weights to update, but also makes the second stage more difficult to learn, since the vector has more elements. Therefore, we set n to as small a value as possible. The Nguyen and Widrow algorithm selected the initial weights for learning the REE, and we then used the final weights of that network as the initial weights for learning each perturbation $T(f_i)$. For the second stage the network had to learn a mapping from $\Re^6$ to $\Re^6$. We thus selected a larger network with n = 12, and again used the Nguyen and Widrow initial conditions. For the quadratic law of motion we used an identical set-up except that for the first stage n = 3. This created a mapping from $\Re^{10}$ to $\Re^{10}$ for the second stage learning, for which selected n = 20.

The results for the simulations were similar. In both cases the network was able to learn the vector mapping. The final sum of squared error was .01 in the linear case and .55 in the quadratic case. The input-output fixed point tests for the quadratic case are shown in Figures 7 and 8. When we submitted the approximation of the fixed point to the network vector mapping, the fixed point was returned, in the sense that the sum of squared errors between the input vector and the output vector was 0.1 in the linear case and 0.01 in the quadratic case.

Since the fixed point in the function space is a fixed point of the network approximation, agents can find the fixed point with an algorithm such as the Gauss-Newton method. In this paper we use a more robust procedure, the Levenberg-Marquandt (LM) method (see e.g. More 1977). The LM algorithm was set with a tolerance of 0.0001 for the maximum difference between the elements of β and the elements of $f(\cdot,\beta)$. With this tolerance, in the local neighborhood of the fixed point, the LM method found the fixed point. Figure 9 shows the graph of the fixed point in the function space found by the LM method for the quadratic case. The final sum of squared errors between the approximation of the fixed point and the fixed point calculated by the LM method was 0.12. We also note that the *stationary state* of the actual REE fixed point is 1, while the stationary state of the fixed point calculated by the LM method was 1.03, indicating that the true economy and the approximated economy possessed essentially the same stationary state price.

4 Conclusions

In this paper we have examined whether or not rational expectations can be learned using incremental learning, intermediate stage learning and two stage learning in a general single good, non-linear self-referential model. We find two main conclusions. First, agents do not learn to be rational using intermediate stage learning. Rather, the sequence of forecast functions generated by the types of learning diverges. Agents can, however, learn to be rational with two stage learning. The key difference is that with intermediate stage learning the TEF changes, while in two stage learning the functional operator T remains fixed. It is this fixed nature of the functional operator which is apparently more important for convergence than in what space (real number or function) the operator works in.

Furthermore, we find that neural networks can be part of a learning algorithm which learns both TEFs and functional mappings. Since neural networks learn by combining transcendental functions in a non-linear manner, they are capable of learning non-linear TEFs and making approximations of functions used for two stage learning. Hence two stage learning with neural networks is a practical, 'hands-on' algorithm for rational learning.

In this study the temporary equilibrium function is first order. In OLG models with several goods, the law of motion can be second order even when the rational expectations forecast function is first order (see e.g. Kehoe and Levine 1984). Analysis of the more general model in which the law of motion possesses higher orders determines that divergence away from the REE with incremental and intermediate stage learning is a facet of the single good, first order economy. In the more general OLG model, incremental learning converges locally to the REE for an open set of economies (Shorish, 1996).

An avenue of future research worth investigating concerns the importance of the two-step forward looking nature of intermediate stage learning. DeCanio (1979) shows that an intermediate stage learning model which is not one-step forward looking converges to an REE, while conversely a multi-step ahead forecast sequence was unstable using incremental learning in Bullard and Duffy (1993). One conclusion from this previous line of research is that there appears to be a close connection between convergence of the learning rule and the model within which the learning rule is embedded. Learning seems to be a context-dependent phenomenon. Exactly which classes of models (in the sense of how forward looking the model is) are learnable with both incremental and intermediate stage learning is a question currently under study.

In addition, we have assumed a representative agent model, where agents always submit identical forecasts. Although this will generally be true at the REE, it is more reasonable to assume that agents possess many different beliefs while learning. For example, agents may have different initial forecast functions, or different beliefs about the temporary equilibrium sequence or about the functional mapping. Other future research would entail a simulation of economies where, for example, agents begin their learning with different types of forecast functions.

Although there is room for extensions to different models, non-linear one-step forward looking models with rational expectations are quite common in the literature. By showing that the two stage learning process converges to the REE, we strengthen the results established in the literature for these models.

5 Appendix: Proofs

5.1 Proof of Theorem 1

Consider the TEF:

$$\int z(p_t, w_t; f(f(p_{t-1}, w_t, \beta), w_{t+1}, \beta))dG = 0 . \qquad 5.1.1$$

(For brevity we omit the dependence of *z* on w_t and of *G* on w_{t+1} in what follows). By Assumption A2, we can restrict our attention to an agent who learns the TEF arbitrarily accurately. Note also that *T* is the implicit TEF function

$$p_t = T(f_\tau) \qquad 5.1.2$$

Hence

$$f_{\tau+1} = p_t^e \in N_\varepsilon(T(f_\tau)) \qquad 5.1.3$$

For ε sufficiently small, then, the convergence properties are identical to the convergence properties of

$$\int z(T(f_\tau); f_\tau \circ f_\tau)dG \equiv 0 . \qquad 5.1.4$$

If the operator T is unstable, the sequence of forecast functions diverges away from the REE.[8] Thus,

$$\left\| T'(f^*) \right\| \geq 1 \Rightarrow \lim_{t\to\infty} \left\| f_\tau - f^* \right\| \neq 0 \qquad 5.1.5$$

By Assumption A1, $T(f_\tau)$ is Fréchet differentiable in f_τ . The Fréchet (equivalently, Gateaux) derivative of *T* is found by perturbing f_τ by a function αh, $\alpha \in \Re, h \in C^1, \|h\| \leq 1$ and then finding the limit of the derivative with respect to (hereafter w.r.t.) α as α approaches zero. We apply the derivative to f_τ in equation 5.1.4:

$$\frac{d}{d\alpha}\int z(T(f_\tau + \alpha h); (f_\tau + \alpha h) \circ (f_\tau + \alpha h))dG \equiv 0 \Rightarrow$$

$$\lim_{\alpha\to 0}\int T'(f_\tau + \alpha h)hz_1(T(f_\tau + \alpha h); (f_\tau + \alpha h) \circ (f_\tau + \alpha h))dG +$$

$$\lim_{\alpha\to 0}\int z_2(T(f_\tau + \alpha h); (f_\tau + \alpha h) \circ (f_\tau + \alpha h))[h(f_\tau + \alpha h) + h \cdot (f_{1,\tau} + \alpha h_1) \circ (f_\tau + \alpha h)]dG = 0 \qquad 5.1.6$$

[8] The reverse is not necessarily true. Additional assumptions would be needed to keep stochastic realizations from bouncing forecast functions *f* outside of the domain of attraction for f^*.

where $z_i, f_{i,\tau}, h_i$ are the derivatives of z, f_τ, and h w.r.t. their i[th] arguments, respectively. Taking the limit as α goes to zero yields a solution for the Fréchet derivative of *T*, T', multiplied by *h*:

$$\lim_{\alpha\to 0}\frac{\partial T}{\partial \alpha} = T'(f_\tau)h = -\frac{\int z_2(T(f_\tau); f_\tau \circ f_\tau)[h(f_\tau) + f_{1,\tau}(f_\tau)h]dG}{\int z_1(T(f_\tau); f_\tau \circ f_\tau)dG} \qquad 5.1.7$$

At the equilibrium point f^*, equation 5.1.7 becomes

$$T'(f^*)h = -\frac{\int z_2(f^*; f^* \circ f^*)[h(f^*) + f_1^*(f^*)h]dG}{\int z_1(f^*; f^* \circ f^*)dG} \qquad 5.1.8$$

We now make the following side calculation. From the definition of the REE forecasting rule, we know that the first order condition is solved:

$$\int z(f^*; f^* \circ f^*)dG \equiv 0 . \qquad 5.1.9$$

Since 5.1.9 is an identity in the price space, we can take the derivative of both sides w.r.t. p_{t-1}, yielding

$$\int z_1(f^*; f^* \circ f^*) + z_2(f^*; f^* \circ f^*)f_1^*(f^*)dG = 0 . \qquad 5.1.10$$

Solving for $f_1^*(f^*)$ and substituting into equation 5.1.8 gives

$$T'(f^*)h = \frac{\int z_2(f^*; f^* \circ f^*)h(f^*)dG}{\int z_1(f^*; f^* \circ f^*)dG} + h . \qquad 5.1.11$$

Now the norm of the derivative of *T* can be found by taking the sup norm of both sides (recall that the underlying arguments p_{t-1} and w_t are drawn from compact sets):

$$\|T'(f^*)\| = \sup_{\|h\|=1}\left\| \frac{\int z_2(f^*; f^* \circ f^*)h(f^*)dG}{\int z_1(f^*; f^* \circ f^*)dG} + h \right\| . \qquad 5.1.12$$

From the Mean Value Theorem, since w_{t+1} has compact support there exists a $\hat{w}_{t+1}$ and constants k_0 and k_1 such that

$$\|T'(f^*)\| = \sup_{\|h\|=1}\left\| \frac{z_2(f^*; f^*(f^*, \hat{w}_{t+1}))(k_1 - k_0)h(f^*, \hat{w}_{t+1})}{\int z_1(f^*; f^* \circ f^*)dG} + h \right\| . \qquad 5.1.13$$

Using the properties of norms, we find

$$\|T'(f^*)\| = \sup_{\|[h(f^*),h]\|=1}\left\| \left[\frac{z_2(f^*; f^*(f^*, \hat{w}_{t+1}))(k_1 - k_0)}{\int z_1(f^*; f^* \circ f^*)dG}, 1 \right] \cdot \begin{bmatrix} h(f^*, \hat{w}_{t+1}) \\ h \end{bmatrix} \right\| \Rightarrow$$

$$\left\| T'\left(f^*\right) \right\| = \sup_{p \in S, w \in \Omega} \left\| \left[\frac{z_2\left(f^*; f^*\left(f^*, \hat{w}_{t+1}\right)\right)\left(k_1 - k_0\right)}{\int z_1\left(f^*; f^* \circ f^*\right) dG}, 1 \right] \right\| \Rightarrow$$

$$\left\| T'\left(f^*\right) \right\| = \sup_{p \in S, w \in \Omega} \left| \frac{z_2\left(f^*; f^*\left(f^*, \hat{w}_{t+1}\right)\right)\left(k_1 - k_0\right)}{\int z_1\left(f^*; f^* \circ f^*\right) dG} \right| + 1 \geq 1 . \qquad 5.1.14$$

The rational expectations equilibrium is locally unstable with incremental or intermediate stage learning.

■

5.2 Proof of Theorem 2

Recall equation 2.1.12, which specifies the evolution of forecast functions:

$$f_{\tau+1} = \Psi\left(T \circ f_\tau; f_\tau\right). \qquad 5.2.1$$

Following Theorem 1, given ε sufficiently small we have

$$\left\| \frac{d\Psi}{df_\tau}\left(f^\bullet\right) \right\| \geq 1 \Rightarrow \lim_{t \to \infty} \left\| f_\tau - f^* \right\| \neq 0 . \qquad 5.2.2$$

By Assumption A1 and the differentiability of Ψ, equation 2.1.12 is Fréchet differentiable w.r.t. f_τ. Again following Theorem 1,

$$\frac{d\Psi}{df_\tau} h = \Psi_1\left(T\left(f_\tau\right); f_\tau\right) T'\left(f_\tau\right) h + \Psi_2\left(T\left(f_\tau\right); f_\tau\right) h , \qquad 5.2.3$$

which when evaluated at the equilibrium yields

$$\left. \frac{d\Psi}{df_\tau} h \right|_{f_\tau = f^*} = \Psi_1\left(f^*; f^*\right) T'\left(f^*\right) h + \Psi_2\left(f^*; f^*\right) h . \qquad 5.2.4$$

Substitution of $T'h$ from equation 5.1.11 from Theorem 1 then gives

$$\left. \frac{d\Psi}{df_\tau} h \right|_{f_\tau = f^*} = \Psi_1\left(f^*; f^*\right) h + \Psi_1\left(f^*; f^*\right) \frac{\int z_2\left(f^*; f^* \circ f^*\right) h\left(f^*\right) dG}{\int z_1\left(f^*; f^* \circ f^*\right) dG} + \Psi_2\left(f^*; f^*\right) h . \qquad 5.2.5$$

We apply the Mean Value Theorem to the above equation to get

$$\left. \frac{d\Psi}{df_\tau} \right|_{f_\tau = f^*} = \left[\Psi_1\left(f^*; f^*\right) \quad \Psi_1\left(f^*; f^*\right) \frac{z_2\left(f^*; f^* \circ f^*\right)\left(k_1 - k_0\right)}{\int z_1\left(f^*; f^* \circ f^*\right) dG} \quad \Psi_2\left(f^*; f^*\right) \right] \begin{bmatrix} h \\ h\left(f^*, \hat{w}_{t+1}\right) \\ h \end{bmatrix} . \qquad 5.2.6$$

Taking the sup norm of both sides yields

$$\left\|\frac{d\Psi}{df_\tau}\right\| = \sup_{\|[h \;\; h(f^*) \;\; h]\|} \left[\Psi_1(f^*;f^*) \quad \Psi_1(f^*;f^*)\frac{z_2(f^*;f^*\circ f^*)(k_1-k_0)}{\int z_1(f^*;f^*\circ f^*)dG} \quad \Psi_2(f^*;f^*)\right]\begin{bmatrix} h \\ h(f^*,\hat{w}_{t+1}) \\ h \end{bmatrix} \Rightarrow$$

$$\left\|\frac{d\Psi}{df_\tau}(f^*)\right\| = \left\|\Psi_1(f^*;f^*)\right\| + \left\|\Psi_1(f^*;f^*)\frac{z_2(f^*;f^*\circ f^*)(k_1-k_0)}{\int z_1(f^*;f^*\circ f^*)dG}\right\| + \left\|\Psi_2(f^*;f^*)\right\|. \qquad 5.2.7$$

As in Theorem 1, we perform a side calculation based on the consistency condition 2.1.12:

$$\Psi(f^*;f^*) \equiv f^* \Rightarrow$$

$$\Psi_1(f^*;f^*) + \Psi_2(f^*;f^*) = 1. \qquad 5.2.8$$

Substituting equation 5.2.8 into 5.2.7 yields

$$\left\|\frac{d\Psi}{df_\tau}(f^*)\right\| = \left\|\Psi_1(f^*;f^*)\right\| + \left\|\Psi_1(f^*;f^*)\frac{z_2(f^*;f^*\circ f^*)(k_1-k_0)}{\int z_1(f^*;f^*\circ f^*)dG}\right\| + \left\|1-\Psi_1(f^*;f^*)\right\|, \qquad 5.2.9$$

and from the triangle inequality we finally have

$$\left\|\frac{d\Psi}{df_\tau}(f^*)\right\| \geq \left\|\Psi_1(f^*;f^*) + 1 - \Psi_1(f^*;f^*)\right\| + \left\|\Psi_1(f^*;f^*)\frac{z_2(f^*;f^*\circ f^*)(k_1-k_0)}{\int z_1(f^*;f^*\circ f^*)dG}\right\| \Rightarrow$$

$$\left\|\frac{d\Psi}{df_\tau}(f^*)\right\| \geq 1 + \left\|\Psi_1(f^*;f^*)\frac{z_2(f^*;f^*\circ f^*)(k_1-k_0)}{\int z_1(f^*;f^*\circ f^*)dG}\right\| \geq 1. \qquad 5.2.10$$

■

5.3 Proof of Proposition 1

From Lemma 1 we know that $\forall f \in C_1^S, \forall \kappa > 0 \exists \beta^{T\circ f} \in B, k(p,\beta^{T\circ f}) \in C_{S\times B}^1$ such that

$$\left\|T\circ f(p) - k(p,\beta^{T\circ f})\right\| < \kappa. \qquad 5.3.1$$

This allows us to define a new mapping $T_k(f) \equiv k(p,\beta^{T\circ f})$. Again from Lemma 1 we know that there is a vector-parameterized approximation to this mapping, i.e. $\forall f \in C_1^S, \forall \nu > 0 \exists k(p,\beta^{T_k\circ f})$ such that

$$\left\|T_k(f) - k(p,\beta^{T_k\circ f})\right\| < \nu. \qquad 5.3.2$$

Then by the triangle inequality and using the inequalities 5.3.1 and 5.3.2 we have

$$\left\|k(p,\beta^{T_k\circ f}) - T(f)\right\| < \left\|k(p,\beta^{T_k\circ f}) - k(p,\beta^{T\circ f})\right\| + \left\|k(p,\beta^{T\circ f}) - T(f)\right\| < \kappa + \nu \equiv \varepsilon.$$

5.3.3

Define $\hat{T}(f) \equiv k\left(f, \beta^{T_k \circ f}\right)$ for brevity. The, since κ and ν are arbitrary, we have that $\forall \varepsilon > 0$,

$$\left\| \hat{T}(f) - T(f) \right\| < \varepsilon . \qquad 5.3.4$$

This concludes the proof of the first part of Proposition 1.

To prove the second part, we first write out the definition of the Fréchet derivative for the operators T and $\hat{T}$:

$$\begin{aligned} \hat{T}'(f)h &= \hat{T}(f+h) - \hat{T}(f) - o(\|h\|) \\ T'(f)h &= T(f+h) - T(f) - o(\|h\|) \end{aligned}. \qquad 5.3.5$$

Taking the sup norm of both sides and subtracting the equations in 5.3.5 yields (ignoring higher order terms)

$$\left\| \|\hat{T}'(f)\| - \|T'(f)\| \right\| = \left| \sup_{\|h\|=1} \left\| \hat{T}'(f+h) - \hat{T}'(f) - o(\|h\|) \right\| - \sup_{\|h\|=1} \left\| T'(f+h) - T'(f) - o(\|h\|) \right\| \right| \leq$$

$$\left| \sup_{\|h\|=1} \left\| \hat{T}'(f+h) - \hat{T}'(f) - o(\|h\|) - T'(f+h) + T'(f) + o(\|h\|) \right\| \right| \leq$$

$$\left| \sup_{\|h\|=1} \left\| \hat{T}'(f+h) - T'(f+h) \right\| + \sup_{\|h\|=1} \left\| \hat{T}'(f) - T'(f) \right\| \right| \leq 2\varepsilon . \qquad 5.3.6$$

■

5.4 Proof of Theorem 3

We omit the first two parts of the proof for brevity—they follow directly from the existence of f^* and from the fact that $\hat{T}$ approximates T arbitrarily closely. For the third part, we first suppose that $\left\| \hat{T}' \right\| < 1$. We know that

$$\left\| k^* - f^* \right\| = \left\| \hat{T}(k^*) - T(f^*) \right\| \equiv \varepsilon^* \Rightarrow$$

$$k^* = f^* + \zeta(p.\beta), \ \|\zeta\| = \varepsilon^* . \qquad 5.4.1$$

From the triangle inequality and Proposition 1 we have

$$\left\| \hat{T}(k^*) - T(f^*) \right\| \leq \left\| \hat{T}(k^*) - \hat{T}(f^*) \right\| + \varepsilon . \qquad 5.4.2$$

Using the definition of the Fréchet derivative (and ignoring higher order terms)

$$\left\| \hat{T}(k^*) - \hat{T}(f^*) \right\| = \left\| \hat{T}(k^*) - \hat{T}(k^* - \zeta) \right\| = \left\| \hat{T}'(k^*)\zeta \right\| < \left\| \hat{T}'(k^*) \right\| \varepsilon^* . \qquad 5.4.3$$

Substituting 5.4.1 and 5.4.3 into 5.4.2 gives

$$\left\|\hat{T}(k^*)-T(f^*)\right\| \le \left\|\hat{T}'(k^*)\right\|\varepsilon^* + \varepsilon \Rightarrow$$

$$\varepsilon^* < \frac{1}{1-\left\|\hat{T}'(k^*)\right\|}. \qquad 5.4.4$$

Now suppose that $\left\|\hat{T}'\right\| > 1$. Since $\hat{T}$ is continuous, there exists a function f_a such that $\hat{T}(f_a) = f^*$. Then by the triangle inequality

$$\left\|k^* - f^*\right\| \le \left\|k^* - f_a\right\| + \left\|f_a - f^*\right\|. \qquad 5.4.5$$

By the Intermediate Value Theorem, there exists a function f_b such that

$$\left\|f_a - f^*\right\|\left\|T'(f_b)\right\| = \left\|T(f_a) - T(f^*)\right\| \Rightarrow$$

$$\left\|f_a - f^*\right\| < \frac{\varepsilon}{\left\|T'(f_b)\right\|}. \qquad 5.4.6$$

Also from the Intermediate Value Theorem there exists a function f_c such that

$$\left\|k^* - f_a\right\|\left\|T'(f_c)\right\| = \left\|\hat{T}(k^*) - \hat{T}(f_a)\right\| = \varepsilon^*. \qquad 5.4.7$$

Substituting equations 5.4.7 and 5.4.6 into 5.4.5 yields

$$\varepsilon^* < \frac{\varepsilon^*}{\left\|\hat{T}'(f_c)\right\|} + \frac{\varepsilon}{\left\|T'(f_b)\right\|} \Rightarrow$$

$$\varepsilon^* \le \varepsilon\left[\frac{\left\|\hat{T}'\right\|}{\left\|T'\right\|\left(\left\|\hat{T}'\right\| - 1\right)}\right]. \qquad 5.4.8$$

■

5.5 Proof of Corollary 1

From Theorem 3 we know that when $\left\|\hat{T}'\right\| > 1$,

$$\varepsilon^* \le \varepsilon\left[\frac{\left\|\hat{T}'\right\|}{\left\|T'\right\|\left(\left\|\hat{T}'\right\| - 1\right)}\right].$$

From this it is evident that if

$$\left\|\hat{T}'\right\| \le \left\|T'\right\|\left[\left\|\hat{T}'\right\| - 1\right], \qquad 5.5.1$$

then $\varepsilon^* \le \varepsilon$. From Proposition 1,

$$\left| \left\| \hat{T}' \right\| - \left\| T' \right\| \right| < 2\varepsilon \Rightarrow \left\| T' \right\| < 2\varepsilon + \left\| \hat{T}' \right\| . \qquad 5.5.2$$

Substitution of the above equation into 5.5.1 then yields

$$\left\| \hat{T}' \right\| \leq \left[\left\| \hat{T}' \right\| + 2\varepsilon \right] \left[\left\| \hat{T}' \right\| - 1 \right]. \qquad 5.5.3$$

Suppose for the moment that 5.5.3 holds with equality—the equation can then be simplified to

$$\left(\left\| \hat{T}' \right\| \right)^2 + (2\varepsilon - 2) \left\| \hat{T}' \right\| - 2\varepsilon = 0 \Rightarrow$$

$$\left\| \hat{T}' \right\|^* = (1 - \varepsilon) + \left(1 + \varepsilon^2\right)^{\frac{1}{2}}, \qquad 5.5.4$$

where $\left\| \hat{T}' \right\|^*$ is the (positive) value of the solution to the above quadratic equation.

By inspection, it is clear that if $\left\| \hat{T}' \right\| > \left\| \hat{T}' \right\|^*$ then equation 5.5.1 obtains, that is if

$$\left\| \hat{T}' \right\| > (1 - \varepsilon) + \left(1 + \varepsilon^2\right)^{\frac{1}{2}},$$

then

$$\varepsilon^* \leq \varepsilon \left[\frac{\left\| \hat{T}' \right\|}{\left\| T' \right\| \left(\left\| \hat{T}' \right\| - 1 \right)} \right] < \varepsilon .$$

■

5.6 Proof of Theorem 4

First define a vector mapping $G : \Re^n \times \Re^n \to \Re$ as

$$G(\beta, \beta_1) \equiv \left\| \hat{T}(k(p, \beta)) - k(p, \beta_1) \right\| . \qquad 5.6.1$$

From Proposition 1 we can infer that there exists a mapping $v : \Re^n \to \Re$ such that $G(\beta, \beta_1) = v(\beta)$, and that

$$v(\beta) = 0 \text{ iff } \hat{T}(k(p, \beta)) = k(p, \beta). \qquad 5.6.2$$

Using the Implicit Function Theorem we know there exists a mapping $\hat{\hat{T}} : \Re^n \to \Re^n$ such that

$$G\left(\beta, \hat{\hat{T}}(\beta) \right) - v(\beta) = 0 . \qquad 5.6.3$$

From this, it is clear that $\hat{\hat{T}}(\beta^*) = \beta^*$ if and only if $\hat{T}(k(p, \beta^*)) = k(p, \beta^*)$.

Since $k\left(p,\beta^{*}\right)$ is a fixed point of $\hat{T}$, by local uniqueness it must be true that $k\left(p,\beta^{*}\right)=k^{*}$, and the properties of Theorem 3 are satisfied.

■

6 Figures

Figures 1 – 6: Divergence of sequence of forecast functions using intermediate stage learning. Solid line = REE, Dashed Line = forecast function. Starred points = training grid of TEF values.

Figure 1

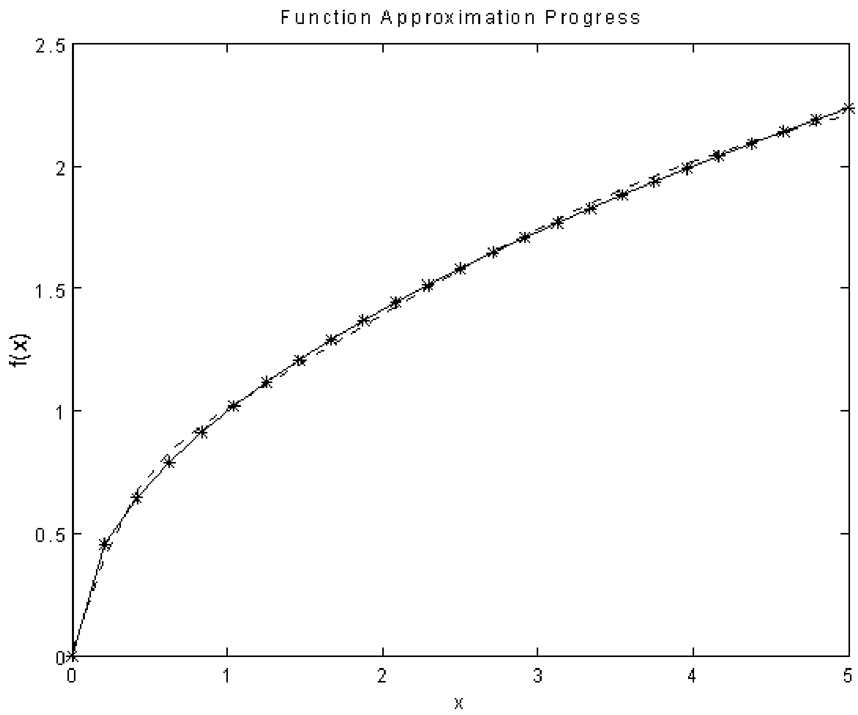


Figure 2

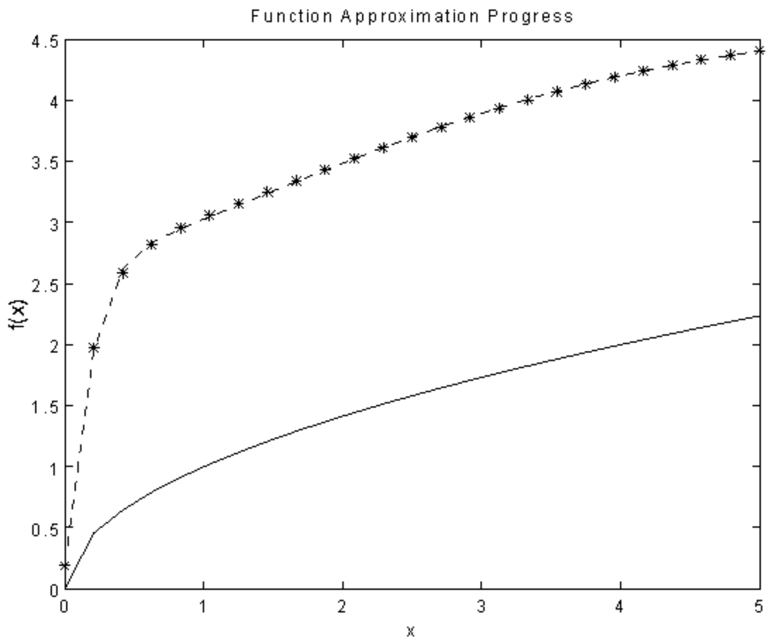


Figure 3

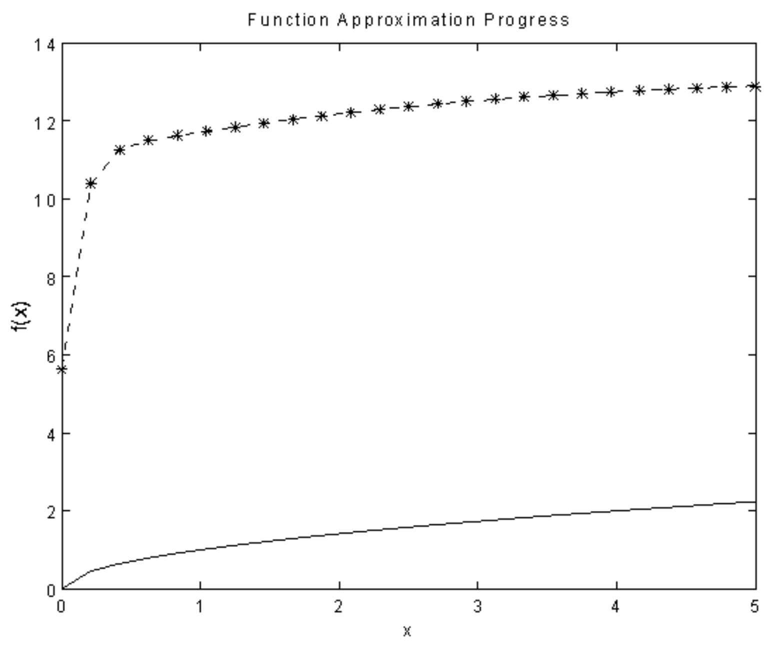


Figure 4

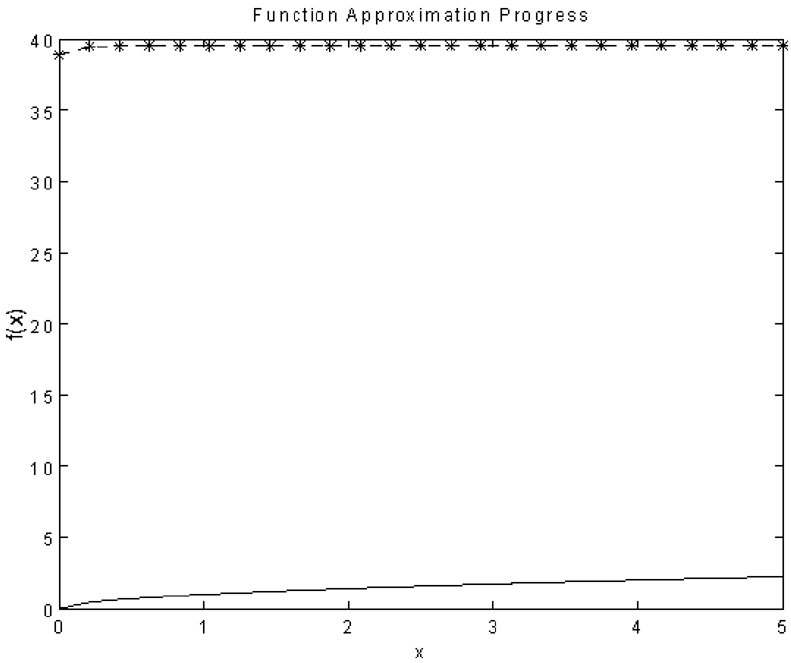


Figure 5

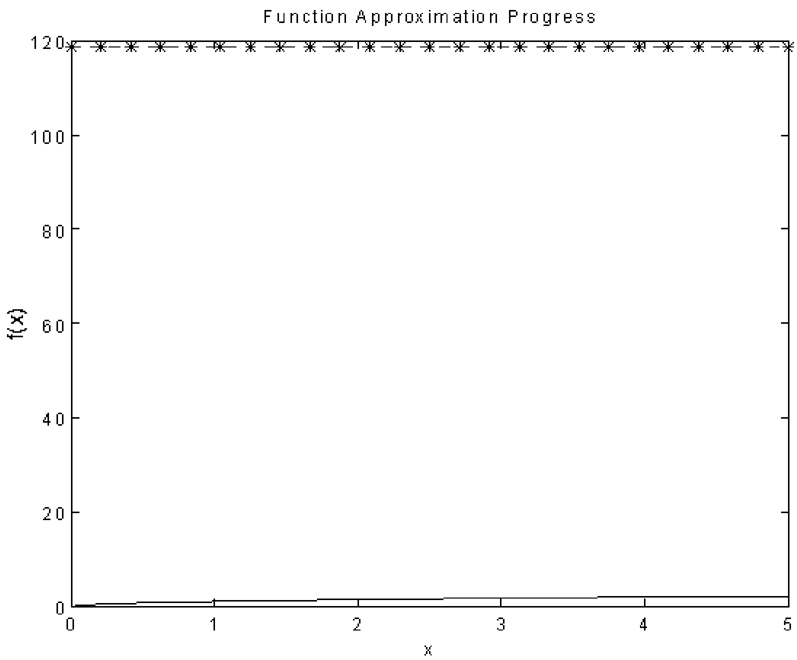


Figure 6

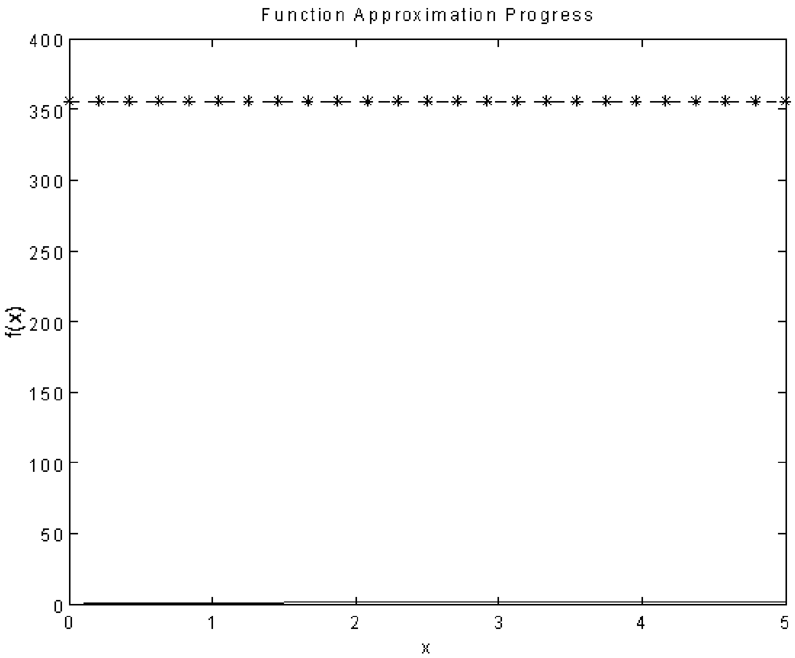


Figure 7. Two Stage Learning: Neural Network Approx. and Actual Fixed Point

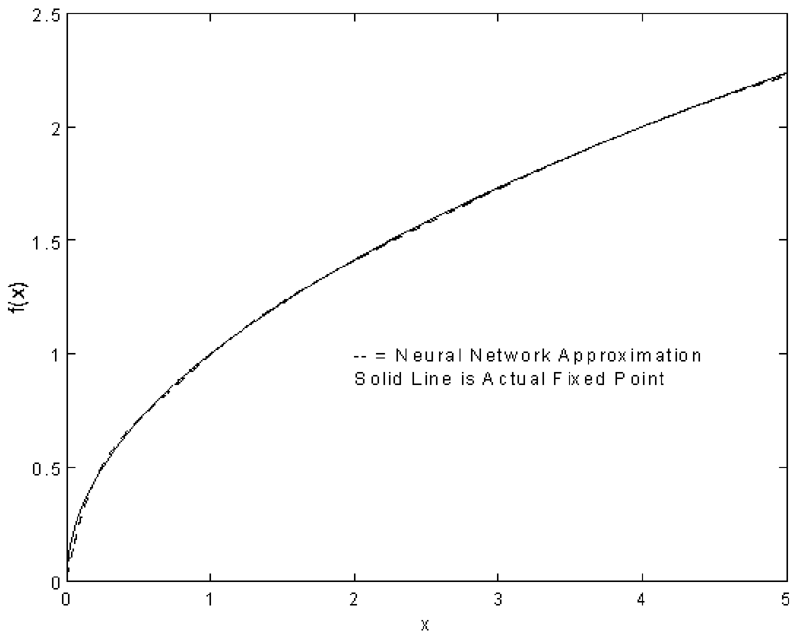


Figure 8. Two Stage Learning: Output of Operator *T* Approximator when Fixed Point Approx. is Input

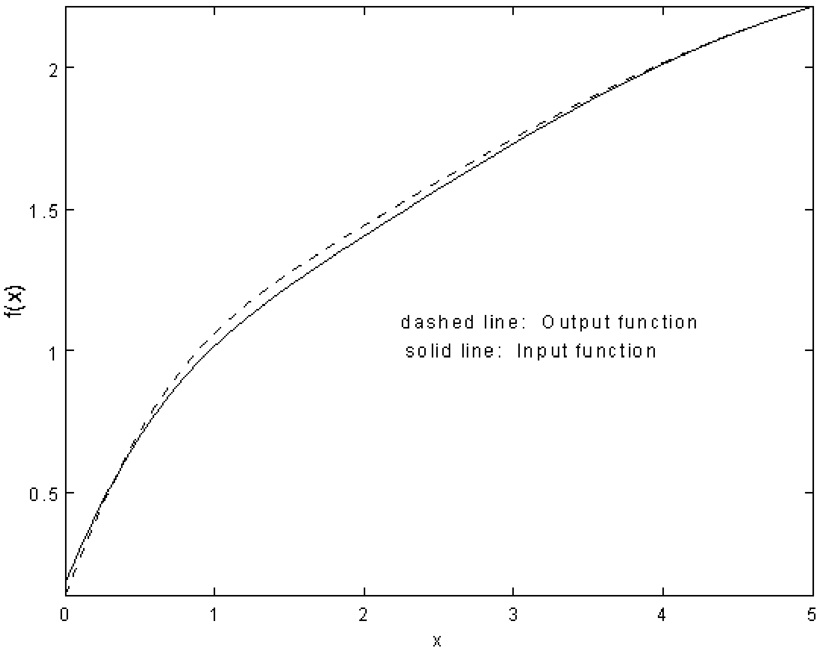
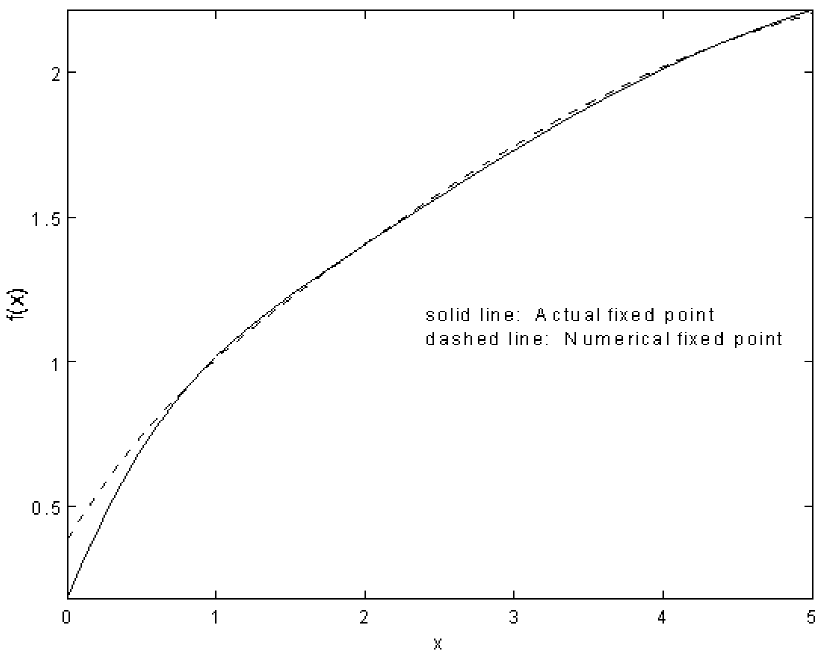


Figure 9. Two Stage Learning: LM Approximation and Actual Fixed Point



7 References

ARIFOVIC, J. (1989) Learning by Genetic Algorithms in Economic Environments, *Santa Fe Institute Working Paper #90-01*.

AZARIADIS, C. (1981) Self Fulfilling Prophecies, *Journal of Economic Theory* 25.

BANSAL, R. AND S. VISWANATHAN. (1993) No Arbitrage and Arbitrage Pricing: A New Approach, *Journal of Finance* 48(4).

BENHABIB, J. AND R. DAY. (1982) A Characterization of Erratic Dynamics in the Overlapping Generations Model, *Journal of Economic Dynamics* 4.

BLUME, L. E. AND D. EASLEY. (1981) Rational Expectations Equilibrium and the Efficient Markets Hypothesis, *Cornell University* .

BOARD, R. (1992) Polynomial Bounded Rationality, *Federal Reserve Board*, Washington D.C.

BRAY, M. AND N. SAVIN. (1986) Rational Expectations Equilibria learning and Model Specification, *Econometrica* 54.

BROCK, W. (1972) On Models of Expectations That Arise From Maximizing Behavior of Economic Agents Over Time, *Journal of Economic Theory* 5.

BULLARD, J. AND J. DUFFY. (1993) Learning in a Large Square Economy, *Federal Reserve Bank of St. Louis and the University of Pittsburgh*.

CASS, D. *ET AL*. (1979) The Role of Money in Supporting the Pareto Optimality of Competitive Equilibrium in Consumption-Loan Type Models, *Journal of Economic Theory* 20(1).

CYERT, R. AND M. DEGROOT. (1974) Rational Expectations and Bayesian Analysis, *Journal of Political Economy* 82.

DECANIO, S. J. (1979) Rational Expectations and Learning from Experience, *Quarterly Journal of Economics* 93.

EASLEY, D. AND N. KIEFER. (1989) Optimal Learning with Endogenous Data, *International Economic Review* 30(4).

EL-GAMAL, M. AND R. SUNDARAM. (1989) Bayesian Economists...Bayesian Agents I: An Alternative Approach to Optimal Learning, *California Institute of Technology*.

EVANS, G. AND G. RAMEY. (1992) Expectation Calculation and Macroecnomic Dynamics, *American Economic Review* 82(1).

FELDMAN, M. (1987) Bayesian Learning and Convergence to Rational Expectations, *Journal of Mathematical Economics* 16.

GALLANT, A. R. AND H. WHITE. (1992) On Learning the Derivatives of an Unknown Mapping With Multilayer Feedforward Networks, *Neural Networks* 5(1).

GRANDMONT, J. M. (1977) Temporary General Equilibrium Theory. *Econometrica* 45.

GRANDMONT, J. M. (1985) On Endogenous Competitive Business Cycles, *Econometrica* 53.

HARP, S. *ET AL*. (1989) Towards the Genetic Synthesis of Neural Networks, *International Conference on Genetic Algorithms* .

HORNIK, K. *ET AL*. (1989) Multi-Layer Feedforward Networks are Universal Approximators, *Technical Report 88-45R, Dept. of Economics, University of California-San Diego.*

HORNIK, K. *ET AL*. (1990) Universal Approximation of an Unknown Mapping and its Derivatives Using Multilayer Feedforward Networks, *Technical Report 89-36R, Dept. of Economics, University of California-San Diego.*

KEHOE, T., AND D.LEVINE. (1984) Regularity in Overlapping Generations Exchange Economies, *Journal of Mathematical Economics* 13.

KURKOVA, V. (1992) Kolmogorov's Theorem and Multilayer Feedforward Networks, *Neural Networks* 5.

LAPEDES, A., AND R. FARBER. (1987) Nonlinear Signal Processing Using Neural Networks: Prediction and System Modeling, *Theoretical Division, Los Alamos National Laboratory.*

LUCAS, R. E. (1986) Adaptive Behavior and Economic Theory, *Journal of Business* 59.

MARCET, A. AND T. SARGENT. (1989a) Least Squares Learning and the Dynamics of Hyperinflation, *In Barnett, W., Geweke, J. and K. Shell (eds.), Economic Complexity: Chaos, Sunspots, Bubbles and Nonlinearity, Proceedings of the Fourth International Symposium in Economic Theory and Econometrics.* Cambridge University Press.

MARCET, A. AND T. SARGENT. (1989b) Convergence of Least Squares Learning Mechanisms in Self-Referential Linear Stochastic Models, *Journal of Economic Theory* 48.

MCCALLUM, B. (1989) New Classical Macroeconomics: A Sympathetic Account, *Scandanavian Journal of Economics* 91.

MCCALLUM, B. (1987) Inflation: Theory and Evidence, *Technical Report 2312, NBER Working Paper*.

MCCLELLAND, J. AND D. RUMELHART. (1988) *Explorations in Parallel Distributed Processing.* MIT Press.

MILLER, G. *ET AL*. (1989) Designing Neural Networks Using Genetic Algorithms, *International Conference on Genetic Algorithms.*

MORE, J. J. (1977) The Levenberg-Marquandt Algorithm: Implementation and Theory, *In Watson, G. (ed.), Numerical Analysis.* Springer-Verlag.

MUTH, J. (1961) Rational Expectations and the Theory of Price Movements, *Econometrica* 29.

NGUYEN, D. AND B. WIDROW. (1990) Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights, *International Joint Conference on Neural Networks* 3.

NYARKO, Y. (1991) The Convergence of Bayesian Belief Hierarchies, *C. V. Starr Center, New York University*.

SCHAFFER, D. *ET AL*. (1990) Using Genetic Search to Exploit the Emergent Behavior of Neural Networks, *Physica D* 42.

SPEAR, S. E. (1988) Existence and Local Uniqueness of Functional Rational Expectations Equilibria in Dynamic Economic Models, *Journal of Economic Theory* 44.

SPEAR, S. E. (1989) Learning Rational Expectations Under Computability Constraints, *Econometrica* 57.

TOWNSEND, R. M. (1983) Forecasting the Forecasts of Others, *Journal of Political Economy* 91(4).

VOGL, T. *ET AL*. (1988) Accelerating the Convergence of the Backpropagation Model, *Biological Cybernetics* 59.

WHITE, H. (1989) Learning in Artificial Neural Networks: A Statistical Perspective, *University of California-San Diego Working Paper.*

WHITE, H. (1992) *Artificial Neural Networks: Approximation and Learning Theory.* Blackwell Publishers.

Working Paper

1998-12 Robert R. Dogonowski: Income Taxation, Imperfect Competition and the Balanced Budget Multiplier.

1998-13 Ebbe Yndgaard: EU Enlargement - Latvia is Ready.

1998-14 Torben M. Andersen: Staggered Wage-Setting and Output Persistence.

1998-15 Henning Bunzel, Bent Jesper Christensen, Niels Haldrup, Svend Hylleberg, Viggo Høst, Peter Jensen og Allan Würtz: Udviklingslinier i Økonometrien.

1998-16 Peter Skott: Wage Formation and the (Non-)Existence of the NAIRU.

1998-17 Svend Hylleberg and Per Baltzer Overgaard: Competition Policy with a Coasian Prior?

1998-18 Lykke Eg Andersen and Osvaldo Nina: Micro-Credit and Group Lending: The Collateral Effect.

1998-19 Torben M. Andersen and Svend Hylleberg: Sources of Persistence in Employment Adjustment - Denmark 1974-1993.

1999-1 Bo Sandemann Rasmussen: Balancing the Budget: Long-Run Adverse Effects of Progressive Taxation.

1999-2 Palle Schelde Andersen, Mark Klau, and Ebbe Yndgaard: Higher Profits and Lower Capital Prices: Is Factor Allocation Optimal?

1999-3 Niels Haldrup and Michael Jansson: Spurious Regression, Cointegration, and Near Cointegration: A Unifying Approach.

1999-4 Christian M. Dahl and Svend Hylleberg: Specifying Nonlinear Econometric Models by Flexible Regression Models and Relative Forecast Performance.

1999-5 David L. Kelly and Jamsheed Shorish: Stability of Functional Rational Expectations Equilibria.