

ADAPTIVE LEARNING IN PRACTICE*

EVA CARCELES-POVEDA, SUNY STONY BROOK

CHRYSSI GIANNITSAROU[†], UNIVERSITY OF CAMBRIDGE AND CEPR

September 18, 2006

ABSTRACT. We analyze practical aspects of implementing adaptive learning in the context of forward looking linear models. We focus on how to set initial conditions for three popular algorithms, namely recursive least squares, stochastic gradient and constant gain learning. We propose three ways of initializing, one that uses randomly generated data, one that is ad-hoc and one that uses an appropriate distribution. We illustrate via standard examples, that the behavior of macroeconomic variables not only depends on the learning algorithm, but on the initial conditions as well. Furthermore, we provide a computing toolbox for analyzing the quantitative properties of dynamic stochastic macroeconomic models under adaptive learning.

KEYWORDS: Adaptive learning, computational methods, short-run dynamics, initial conditions.

JEL CLASSIFICATION: C63, D83, E10

*We thank Seppo Honkapohja and Jesus Fernandez Villaverde for helpful suggestions, as well as Damjan Pfajfar for research assistance. For comments, we are grateful to conference and seminar participants of SITE 2004, CEF 2004, EEA-ESEM 2004 and Heriot-Watt University. This project has been supported by the ESRC award RES-000-23-1152.

[†]Corresponding author. *Address:* Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, UK. *E-mail:* cg349@cam.ac.uk. *Telephone Number:* +44 1223 335250. *Web:* <http://www.econ.cam.ac.uk/faculty/Giannitsarou/index.html>

1. INTRODUCTION

By now, there exists an extensive and well established literature on identifying and studying the asymptotic properties of adaptive learning algorithms in macroeconomic models.¹ However it was not until very recently that adaptive learning has gained popularity in applied macroeconomics, especially in dynamic general equilibrium settings. For example, recent literature has used adaptive learning to study topics like the evolution of US inflation and the importance of expectations for its determination, the effects of monetary policy on macroeconomic variables, hyperinflations, business cycle fluctuations, asset prices, structural changes and policy reforms (see for example Sargent, 1999, Cho, Williams and Sargent, 2002, Bullard and Cho, 2005, Marcet and Nicolini, 2003, Williams, 2003, Milani, 2005a, 2005b, 2005c, Orphanides and Williams, 2005a, 2005b, Bullard and Eusepi, 2005, Carceles-Poveda and Giannitsarou, 2006 and Giannitsarou, 2006).

The aim of our paper is to contribute to this rapidly expanding quantitative and empirical literature by (i) providing a unified and self-contained framework that can be used to analyze the learning dynamics of a wide class of dynamic stochastic macroeconomic models numerically and (ii) illustrating the importance of initial conditions for different learning algorithms for short and medium time horizons. Short horizons are of particular relevance for macroeconomics, since typical time series are rather short. Moreover, short run learning dynamics are important when considering the effects of structural changes, since they allow for gradual adjustment towards a new equilibrium. As shown in Bullard and Eusepi (2005) and Giannitsarou (2006), short run learning dynamics can be very different from the corresponding rational expectations dynamics.

We study two general reduced forms. The first is a forward looking model with a lagged endogenous variable. This particular reduced form often arises in macroeconomics, as it may result from dynamic nonlinear general equilibrium models that can be log-linearized around a steady state, such as the stochastic growth model and the basic new Keynesian framework. The second reduced form is a purely forward looking model, in the sense that it does not include a lag of the endogenous state variable. We provide two examples of such models. The first, corresponding to the reduced form with a lag, is the standard stochastic growth model with inelastic labor supply. The second example is the Lucas tree model, which has a purely forward looking reduced form.

First, we explain how to rewrite all the examples in the generic reduced form once they are log-linearized around their steady states.² Once a model is written in reduced form, we then explain how to get the rational expectations equilibria, by deriving minimum state variable (MSV) solutions using the method of undetermined coefficients. Here, we focus on models with a unique stationary solution.

Following this, we turn to a brief discussion of general adaptive learning algorithms and their asymptotic properties. We focus on three different algorithms, recursive least squares,

¹For a comprehensive analysis of the asymptotic properties of adaptive learning algorithms in economics, see Evans and Honkapohja (2001).

²Note that, while we focus on a univariate specification, our work can easily be extended to a multivariate framework.

stochastic gradient and constant gain learning, for which we briefly review the conditions that must hold in order for them to converge to the unique rational expectations equilibrium. For the least squares and constant gain algorithms, we provide a derivation of their recursive forms. As part of these derivations, we point out some technical complications that arise when ensuring that the recursion is well defined.

Next we discuss how to initialize the recursion and propose three possible ways of setting the initial conditions. One is based on randomly generated data before the recursion begins, the second just sets ad-hoc initial conditions and the last is a variant of the first that initializes using an appropriate distribution of the coefficient to be estimated. We argue that each of these ways of initializing the algorithms is more appropriate depending on the situation the researcher wishes to study. The first way of initializing is reasonable if one assumes that it takes some time to collect data; the ad-hoc method could be useful when studying structural changes or when the initial condition can be estimated from data; finally, the distribution method could be suitable in a context with heterogeneous initial priors, since the algorithm would start from some average beliefs.

Next, we discuss important differences between the three learning algorithms, as well as their advantages and disadvantages given the context within which they are used. For example, stochastic gradient learning is relatively simple, and there are no technical complications related to the recursion when initializing the algorithm. However it is not clear that it should be preferred to recursive least squares, since it is inefficient (in the econometric sense) and its speed of convergence depends crucially on the size of the variance of the exogenous shocks. Moreover, the estimated coefficients under constant gain learning do not converge to a single point but to a distribution around the rational expectations value of the estimate. Whereas this represents a bigger departure from rational expectations, it might be more useful than the other algorithms for explaining particular stylized facts.

Finally, we illustrate in detail how to numerically implement learning through the two examples and provide comparisons between possible outcomes for different initial conditions and different algorithms. These are done graphically, but also by providing statistics such as relative deviations and correlations of the variables of interest, which are typically used in the macroeconomic literature. In general, we find that the initial conditions may matter a lot for the evolution of the different variables, especially in models with a lagged endogenous variable, where the rational expectations equilibrium is often very close to the stationarity bound, i.e. close to one. Different algorithms also matter for the dynamics but appear to be less important for the evolution of the variables than the initial values.

The paper is accompanied by a programming toolbox which implements all these steps.³ The structure of the toolbox, as well as all the codes, are explained in detail in a manual accompanying the toolbox. The main set of codes is written in a general way so that it can be used for any model satisfying our reduced form models. We also provide model specific codes for analyzing the examples presented in the paper.

³The codes are written for Matlab and Mathematica and are available at <http://www.econ.cam.ac.uk/research/learning/>. Updates for the toolbox will be provided at the same website.

The paper is organized as follows. The next section presents the two reduced form models that we study throughout the paper and provides examples of models that can be written in such reduced forms. Section 3 reviews how to obtain the rational expectations equilibria for such models. Section 4 provides a brief general discussion of adaptive learning in the context of linear models. The next three sections focus on the three algorithms of interest. Section 8 presents numerical results based on the examples we introduce in section 2 and finally section 9 concludes.

2. THE MODELS

In this section we describe the two reduced form forward looking models of interest. The first includes a lag of the endogenous variable, while the second is purely forward looking. Following the two reduced form models, we present two examples of well known economies that can be written in this forms.⁴

2.1. Forward looking reduced form with a lagged endogenous variable (FLL).

We first consider univariate forward looking models with a lagged endogenous variable, which have a reduced form that is given by

$$k_t = a_1 E_t k_{t+1} + a_2 k_{t-1} + b z_t, \quad (1)$$

$$z_t = \rho z_{t-1} + \varepsilon_t, \quad (2)$$

where a_1 , a_2 and b are constants that depend on the model parameters, k_t is the endogenous variable of interest and z_t is an exogenous shock, assumed to follow an AR(1) process with an $iid(0, \sigma^2)$ disturbance ε_t . Throughout the paper, we will assume that the stochastic process is completely known to the agents, i.e. the agents know that z_t follows an AR(1) process and the parameters ρ and σ^2 are also known. This reduced form may derive from a linear model or may arise from the log-linearization around a steady state of a nonlinear model. The original model may also be described by other endogenous (non-state) variables in the model. Such variables can in general be rewritten in terms of the state variables only. For example, for any such variable q_t , we can write

$$q_t = \gamma_{q1} k_t + \gamma_{q2} k_{t-1} + \gamma_{q3} z_t + \gamma_{q4} E_t k_{t+1}, \quad (3)$$

where the coefficients γ_{qi} depend on the parameters of the model.

2.2. Purely forward looking reduced form (PFL). The second reduced form we consider is a special case of (1) - (2) where $a_2 = 0$, so that it does not contain the lag of k_t . The reduced form model is given by

$$k_t = a_1 E_t k_{t+1} + b z_t, \quad (4)$$

$$z_t = \rho z_{t-1} + \varepsilon_t, \quad (5)$$

⁴The accompanying toolbox also contains a third example which is the RBC model with capital and labor income taxes, as in Giannitsarou (2006).

where a_1 and b are constants that depend on the model parameters, k_t is the endogenous variable of interest and z_t is an exogenous shock, assumed to follow an AR(1) process with an $iid(0, \sigma^2)$ disturbance ε_t . Once again, we assume that the parameters characterizing the exogenous process z_t are known. Furthermore, other variables of the original model again take the form

$$q_t = \gamma_{q1}k_t + \gamma_{q3}z_t + \gamma_{q4}E_t k_{t+1}, \quad (6)$$

where the coefficients γ_{qi} depend on the parameters of the model. Since this reduced form does not contain a lag of the variable of interest, the rest of the variables do not depend on k_{t-1} either.

2.3. Examples. We now illustrate the general procedure for deriving the reduced form of nonlinear models via two standard examples: the simple stochastic growth model and the Lucas tree model. After presenting the models, we summarize the necessary equilibrium conditions, the steady state values and the log-linearized system of equilibrium equations, from which we derive the reduced form. The first example fits into the FLL reduced form. The second example fits into the PFL reduced form. Other models that can be written in these reduced forms are variations of the real business cycle model, as well as simple variations of the new Keynesian model.

The general procedure for finding the reduced form of a nonlinear model involves the following steps: (i) determine the equilibrium conditions, such as the constraints and the first order conditions (ii) find the steady state of the model as a function of the parameters, (iii) log-linearize the system of equilibrium conditions to make the equations approximately linear in the log-deviations from the steady state, (iv) reduce the log-linearized system of equilibrium conditions to a system of equations that expresses the current endogenous state variables of interest as linear functions of the exogenous states, as well as one-period-leads and one-period-lags of the endogenous states. Since steps (i)-(iii) have been extensively discussed in the literature by authors like McCallum (1983), Campbell (1994) or Uhlig (1999), we concentrate on step (iv), which we derive for our examples.

In principle, the reduced form of a nonlinear model can be obtained with paper and pencil, but it might be very tedious to do it by hand when the model becomes more complicated. For this reason, we illustrate the procedure by deriving the reduced form coefficients by hand for our first example, and we also provide model specific Mathematica notebooks with the derivation for other examples.⁵ Since the models are relatively standard and the main aim of the section is to illustrate how to calculate the reduced form coefficients and the set of rational expectations equilibria, we focus on the mathematical description without providing a detailed description of the models.

⁵The Mathematica files can be easily modified by the user, to accommodate for a variety of models.

Example 1: The Stochastic Growth Model. The social planner solves the following problem of the representative household:

$$U = \max_{C_t, K_t} E_0 \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma} - 1}{1-\gamma} \quad (7)$$

$$\text{s.t. } C_t + K_t = (1-\delta)K_{t-1} + Y_t, \quad (8)$$

$$Y_t = Z_t K_{t-1}^\alpha, \quad (9)$$

$$\log Z_t = \rho \log Z_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2), \quad (10)$$

where C_t is consumption, K_t is the aggregate capital stock, Y_t is production and Z_t is the total factor productivity. Furthermore, $\beta \in (0, 1)$, $\alpha \in (0, 1)$, $\delta \in (0, 1)$, γ , ρ , and σ , are parameters representing the household discount factor, the capital income share, the capital depreciation, the household risk aversion, the shock persistence and the shock innovation variance respectively. After taking the first order conditions, the system of equilibrium conditions can be summarized by

$$1 = \beta E_t \left[\left(\frac{C_t}{C_{t+1}} \right)^\gamma (\alpha Z_{t+1} K_t^{\alpha-1} + (1-\delta)) \right], \quad (11)$$

$$C_t = (1-\delta)K_{t-1} + Z_t K_{t-1}^\alpha - K_t, \quad (12)$$

$$Y_t = Z_t K_{t-1}^\alpha, \quad (13)$$

$$\log Z_t = \rho \log Z_{t-1} + \varepsilon_t. \quad (14)$$

Next, let the log-deviation of any variable Q_t from its steady state value \bar{Q} be $q_t = \log(Q_t) - \log(\bar{Q})$. Then, the log-linearized conditions are:⁶

$$0 = E_t [\gamma(c_t - c_{t+1}) - \alpha\beta\theta(1-\alpha)k_t + \alpha\beta\theta z_{t+1}], \quad (15)$$

$$c_t = \frac{\bar{K}\bar{R}}{\bar{C}}k_{t-1} + \frac{\bar{Y}}{\bar{C}}z_t - \frac{\bar{K}}{\bar{C}}k_t, \quad (16)$$

$$y_t = z_t + \alpha k_{t-1}, \quad (17)$$

$$z_t = \rho z_{t-1} + \varepsilon_t. \quad (18)$$

where $\theta = (1 - \beta + \delta\beta)/\alpha\beta$.

Once we have the log-linearized system of equations, we can obtain the reduced form of the model by eliminating all the variables apart from the states from the Euler equation. In the present example, we can use the resource constraint and the law of motion of the exogenous shock to substitute for c_t , c_{t+1} and z_{t+1} in the Euler equation until it is a function of the exogenous state variable z_t , the endogenous state k_t , its expected future value $E_t k_{t+1}$ and one lag k_{t-1} . Note that, to eliminate z_{t+1} from the Euler equations, we use that $E_t z_{t+1} = \rho z_t$.

⁶For completeness, the steady state equations for the examples are provided in appendix A. The appendix also contains the derivation of the reduced form for the first example. In addition, the Mathematica files `example_1.nb` and `example_2.nb` contain details of the reduced form coefficient derivations for the two examples. In addition, the file `example_3.nb` contains the same derivations for a variation of the standard RBC model with income taxes.

We then obtain:

$$\left(\frac{\gamma\bar{K} + \gamma\bar{K}\bar{R} + \alpha\beta\theta(1-\alpha)\bar{C}}{\bar{C}}\right)k_t = \frac{\gamma\bar{K}}{\bar{C}}E_t k_{t+1} + \frac{\gamma\bar{K}\bar{R}}{\bar{C}}k_{t-1} + \left(\frac{\gamma\bar{Y} - \gamma\rho\bar{Y} + \alpha\beta\theta\rho\bar{C}}{\bar{C}}\right)z_t. \quad (19)$$

Finally, isolating k_t , we get the reduced form of the model:⁷

$$k_t = a_1 E_t k_{t+1} + a_2 k_{t-1} + b z_t, \quad (20)$$

$$z_t = \rho z_{t-1} + \varepsilon_t, \quad (21)$$

where

$$a_1 = \frac{\gamma\bar{K}}{(\gamma\bar{K} + \gamma\bar{K}\bar{R} + \alpha\beta\theta(1-\alpha)\bar{C})}, \quad (22)$$

$$a_2 = \frac{\gamma\bar{K}\bar{R}}{(\gamma\bar{K} + \gamma\bar{K}\bar{R} + \alpha\beta\theta(1-\alpha)\bar{C})}, \quad (23)$$

$$b = \frac{(\gamma\bar{Y} - \gamma\rho\bar{Y} + \alpha\beta\theta\rho\bar{C})}{(\gamma\bar{K} + \gamma\bar{K}\bar{R} + \alpha\beta\theta(1-\alpha)\bar{C})}. \quad (24)$$

Last, the rest of the endogenous variables of the model, i.e. output and consumption can be written in the form of equation (3), with

$$\gamma_{y1} = 0, \quad \gamma_{y2} = \alpha, \quad \gamma_{y3} = 1, \quad \gamma_{y4} = 0, \quad (25)$$

$$\gamma_{c1} = -\bar{K}/\bar{C}, \quad \gamma_{c2} = -\bar{K}\bar{R}/\bar{C}, \quad \gamma_{c3} = \bar{Y}/\bar{C}, \quad \gamma_{c4} = 0. \quad (26)$$

Example 2: The Lucas tree model. In the Lucas tree model there is no endogenous production, therefore in equilibrium $C_t = D_t$, where D_t is the dividend of the tree. Further, the price of the tree is equal to:

$$P_t = \beta E_t \left(\frac{D_{t+1}}{D_t}\right)^{-\gamma} (D_{t+1} + P_{t+1}), \quad (27)$$

where the exogenous dividend process is given by⁸

$$\log D_t = \rho \log D_{t-1} + \varepsilon_t \quad (28)$$

and $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$. The log-linear system of the equations that describe the model is:

$$p_t = \beta E_t p_{t+1} + (1 - \beta - \gamma) E_t d_{t+1} + \gamma d_t, \quad (29)$$

$$c_t = d_t, \quad (30)$$

$$d_{t+1} = \rho d_t + \varepsilon_{t+1}. \quad (31)$$

⁷When calculating the reduced form, it is important to note that one might encounter problems of invertibility, since some of the terms we have to divide with, such as the numerator multiplying k_t , might be equal to zero. While this problem does not arise in the present examples, it might be the case in the presence of multiplicity or indeterminacy of equilibria, an issue which is outside the scope of the present paper.

⁸In the following specification for the dividend process we have implicitly made the assumption that the steady state for the dividend is $\bar{D} = 1$, implying that $\bar{C} = 1$ and $\bar{P} = \frac{\beta}{1-\beta}$.

Finally, we can write the model in the form of (4) - (5) as follows

$$p_t = a_1 E_t p_{t+1} + b d_t, \quad (32)$$

$$d_t = \rho d_{t-1} + \varepsilon_t. \quad (33)$$

where $a_1 = \beta$ and $b = (1 - \beta - \gamma)\rho + \gamma$. The other variable of interest, c_t is written as $c_t = d_t$ so that $\gamma_{c1} = \gamma_{c4} = 0$ and $\gamma_{c3} = 1$.

3. RATIONAL EXPECTATIONS

To find the rational expectations equilibria of the reduced form models, we use the method of undetermined coefficients, explained in detail in McCallum (1983), Campbell (1994) and Uhlig (1997). We look at each reduced form in turn.

3.1. FLL. We first conjecture that the solution for the endogenous state variable of interest is a function of the state variables of the model k_{t-1} and z_t , i.e., the conjectured solution for k_t is of the form

$$k_t = \eta_{kk} k_{t-1} + \eta_{kz} z_t, \quad (34)$$

where η_{kk} and η_{kz} are the elasticities of k_t with respect to the two state variables. Substituting for $z_t = \rho z_{t-1} + \varepsilon_t$, we obtain

$$k_t = \phi_k k_{t-1} + \phi_z z_{t-1} + \eta_t, \quad (35)$$

where $\phi_k = \eta_{kk}$, $\phi_z = \eta_{kz}\rho$, $\eta_t = \phi_z \varepsilon_t / \rho$ and ϕ_k and ϕ_z are the two undetermined coefficients. Using the previous equation to substitute for the expectation term in the reduced form, we get

$$k_t = a_1 (\phi_k k_t + \phi_z z_t) + a_2 k_{t-1} + b z_t. \quad (36)$$

Rearranging terms and using again the fact that $z_t = \rho z_{t-1} + \varepsilon_t$, the previous equation can be rewritten as

$$(1 - a_1 \phi_k) k_t = a_2 k_{t-1} + (a_1 \phi_z + b) \rho z_{t-1} + (a_1 \phi_z + b) \varepsilon_t. \quad (37)$$

Therefore,

$$k_t = \frac{a_2}{1 - a_1 \phi_k} k_{t-1} + \frac{(a_1 \phi_z + b) \rho}{1 - a_1 \phi_k} z_{t-1} + \frac{a_1 \phi_z + b}{1 - a_1 \phi_k} \varepsilon_t. \quad (38)$$

Finally, since the coefficients of k_{t-1} and z_{t-1} in (35) and (38) have to be the same, we must have that

$$\phi_k = \frac{a_2}{1 - a_1 \phi_k} \quad \text{and} \quad \phi_z = \frac{(a_1 \phi_z + b) \rho}{1 - a_1 \phi_k}, \quad (39)$$

which in turn gives the solutions

$$\bar{\phi}_k = \frac{1}{2a_1} (1 \pm \sqrt{1 - 4a_1 a_2}), \quad (40)$$

$$\bar{\phi}_z = \frac{b\rho}{1 - a_1(\rho + \bar{\phi}_k)}. \quad (41)$$

Clearly, the reduced form model has two solutions $\bar{\phi} = (\bar{\phi}_k, \bar{\phi}_z)$, corresponding to the so-called minimum state variable (MSV) solutions.⁹ Moreover, it is known that the model is regular, i.e. it has a unique stationary solution if and only if $|a_1 + a_2| < 1$ (see Evans and Honkapohja, 2001). In general, if the reduced form derives from the log-linearization of a model similar to our examples, we will have saddle-path stability or a unique stationary solution. If the model yields two stationary solutions, then the model has also non-fundamental solutions, such as sunspot solutions. Here, we only study cases with a unique stationary solution.

3.2. PFL. The rational expectations MSV solution can be obtained by conjecturing that

$$p_t = \phi d_{t-1} + \eta_t, \quad (42)$$

where η_t is some white noise shock. Following an analogous procedure as for the FLL, we must have that

$$\bar{\phi} = (a_1 \bar{\phi} + b) \rho, \quad (43)$$

so that the MSV solution is

$$\bar{\phi} = \frac{b\rho}{1 - a_1\rho}. \quad (44)$$

The solution is unique. It is finite as long as $a_1\rho \neq 1$.

4. ADAPTIVE LEARNING

In this section we provide a brief background on adaptive learning that we will be using throughout the paper. We assume that agents do not form expectations rationally, but instead learn or use estimates of the coefficients $\bar{\phi}$, which they update in every period by employing some econometric technique. Before presenting the general learning algorithm, it is important to note that the forward looking nature of the reduced form models leads to a problem of simultaneity, in the sense that the solution k_t and the coefficients ϕ_t would have to be determined simultaneously if households used ϕ_t to form the period t expectations. To overcome this problem, we assume that agents forecast k_{t+1} using their estimates from the previous period ϕ_{t-1} .¹⁰

For the FLL reduced form, this assumption means that agents form expectations according to¹¹

$$E_t^* k_{t+1} = \phi_{k,t-1} k_t + \phi_{z,t-1} z_t. \quad (45)$$

⁹Given the reduced form coefficients a_1 , a_2 and b , the MSV solutions for our examples can be calculated either by using the Mathematica notebooks `example_1.nb` and `example_2.nb` or by using the Matlab file `solution.m`. For a definition and discussion of MSV solutions, see Evans and Honkapohja (2001). There are some variations in the definitions of an MSV solution in the literature (e.g. see original definition of McCallum, 1983). However, they all boil down to the same principle, which is solutions that are determined by the smallest possible number of state variables (and lags of those).

¹⁰Alternatively, one may assume that k_t is not included in the information set when forming expectations, i.e. expectations are formed using data up to $t - 1$ (see Evans and Honkapohja, 2001).

¹¹We have not included a constant in the expectations formation, as our implicit assumption is that agents are aware that they are estimating deviations of a variable from its steady state. However, an intercept may be easily added to the formulation of the problem without altering the analysis in any significant way.

Inserting this into the reduced form and substituting for $z_t = \rho z_{t-1} + \varepsilon_t$, we obtain

$$\begin{aligned}
k_t &= a_1 (\phi_{k,t-1} k_t + \phi_{z,t-1} z_t) + a_2 k_{t-1} + b z_t \iff \\
(1 - a_1 \phi_{k,t-1}) k_t &= a_2 k_{t-1} + (a_1 \phi_{z,t-1} + b) z_t \iff \\
k_t &= \frac{a_2}{1 - a_1 \phi_{k,t-1}} k_{t-1} + \frac{a_1 \phi_{z,t-1} + b}{1 - a_1 \phi_{k,t-1}} z_t \iff \\
k_t &= \frac{a_2}{1 - a_1 \phi_{k,t-1}} k_{t-1} + \frac{(a_1 \phi_{z,t-1} + b) \rho}{1 - a_1 \phi_{k,t-1}} z_{t-1} + \frac{a_1 \phi_{z,t-1} + b}{1 - a_1 \phi_{k,t-1}} \varepsilon_t \iff \\
k_t &= T_1(\phi_{t-1}) k_{t-1} + T_2(\phi_{t-1}) z_{t-1} + V(\phi_{t-1}) \varepsilon_t, \tag{46}
\end{aligned}$$

where

$$T_1(\phi_{t-1}) = \frac{a_2}{1 - a_1 \phi_{k,t-1}}, \tag{47}$$

$$T_2(\phi_{t-1}) = \frac{(a_1 \phi_{z,t-1} + b) \rho}{1 - a_1 \phi_{k,t-1}}, \tag{48}$$

$$V(\phi_{t-1}) = \frac{a_1 \phi_{z,t-1} + b}{1 - a_1 \phi_{k,t-1}} = \frac{T_2(\phi_{t-1})}{\rho}. \tag{49}$$

Given the previous period estimates ϕ_{t-1} , the expression (46) is the actual law of motion (ALM) for the endogenous state variable k_t . The assumed expectations formation (45) is consistent with the realized k_t , in the sense that, if households form expectations as above, the realized k_t depends on the relevant state variables and it is therefore not misspecified. Finally, if the model contains some other endogenous variable determined as in (3), the actual law of motion of the variable is given by

$$\begin{aligned}
q_t &= [\gamma_{q1} T_1(\phi_{t-1}) + \gamma_{q2} + \gamma_{q4} \phi_{k,t-1} T_1(\phi_{t-1})] k_{t-1} + \\
&\quad [\gamma_{q1} V(\phi_{t-1}) + \gamma_{q3} + \gamma_{q4} \phi_{k,t-1} V(\phi_{t-1}) + \gamma_{q4} \phi_{z,t-1}] z_t. \tag{50}
\end{aligned}$$

For the PFL reduced form, we assume that expectations are formed according to

$$E_t^* k_{t+1} = \phi_{t-1} z_t \tag{51}$$

and with similar arguments we obtain the ALM of k_t ,

$$k_t = T(\phi_{t-1}) z_{t-1} + V(\phi_{t-1}) \varepsilon_t, \tag{52}$$

where

$$T(\phi_{t-1}) = (a_1 \phi_{t-1} + b) \rho, \tag{53}$$

$$V(\phi_{t-1}) = a_1 \phi_{t-1} + b = \frac{T(\phi_{t-1})}{\rho}. \tag{54}$$

Last, the rest of the variables of the model can be expressed as

$$q_t = [\gamma_{q1}V(\phi_{t-1}) + \gamma_{q3} + \gamma_{q4}\phi_{t-1}] z_t. \quad (55)$$

The assumption that agents use ϕ_{t-1} to form expectations about tomorrow's k_{t+1} is based on the following sequence of events during a period: at the beginning of a period, the current state variables are realized (implied by the ALM). These are then used to update the estimates. Finally, they are also used in combination with previous period's estimates to form expectations about the endogenous state variable in the next period. The time line of the learning process is summarized in Figure 1.

< FIGURE 1 ABOUT HERE >

Next, note that a general adaptive learning algorithm can be represented by

$$\theta_t = \theta_{t-1} + g_t Q(t, \theta_{t-1}, x_{t-1}), \quad (56)$$

where θ is a vector of parameters to be estimated, i.e. ϕ and other auxiliary parameters, x is the vector of state variables (which will in general depend on the estimated parameters θ), g_t is the gain sequence and Q determines the updating based on the forecasting error. The gain sequence must be strictly positive and it may be specified as decreasing or constant over time. As we will see in the next sections, the asymptotic results differ depending on the specification. The function Q will typically be a combination of the mappings T and V evaluated at θ_{t-1} .

Given that certain assumptions are satisfied (see Evans and Honkapohja (2001)), the general algorithm (56) can be associated with the ordinary differential equation (ODE)

$$\frac{d\theta}{d\tau} = h(\theta), \quad (57)$$

where $h(\theta) = \lim_{t \rightarrow \infty} E [Q(t, \theta, x_t(\theta))]$. If the ODE has an equilibrium point θ^* which is locally asymptotically stable, then the algorithm converges to θ^* . If θ^* is not an equilibrium point, or if it is not a locally asymptotically stable equilibrium point of the ODE, then the algorithm converges to θ^* with probability zero. If the ODE method can be applied, then the local asymptotic stability of an equilibrium θ^* (and hence convergence of the learning algorithm to it) is determined by the local asymptotic stability of the associated ODE, which in turn is determined by the stability of the Jacobian matrix $J(\theta^*) = D_\theta h(\theta^*)$. Therefore, the conditions required for convergence and stability of the learning algorithm, henceforth *stability conditions*, are derived by imposing that $J(\theta^*)$ is a stable matrix. More detailed discussions on the asymptotic properties of specific algorithms are presented in the next few sections.

5. RECURSIVE LEAST SQUARES (RLS)

In this section, we turn to the recursive least squares algorithm, which is the most widely used algorithm in the learning literature, since it is an intuitive and reasonable estimation technique if the agents in the model are viewed as econometricians. First, we derive the recursive least squares learning algorithm from the usual OLS expression for the particular reduced forms at hand. Although this is a tedious but straightforward exercise (see, for example, Ljung and Söderström (1986)), it is useful to do the derivation in order to clarify some issues regarding the determination of initial conditions and the early periods of estimation.

At the beginning of period t , the vector of state variables x_t is realized, based on the actual law of motion. For the FLL reduced form, the state variables are $x_t = (k_t, z_t)'$ and they are given by

$$x_t = \begin{pmatrix} k_t \\ z_t \end{pmatrix} = \begin{pmatrix} T_1(\phi_{t-1}) & T_2(\phi_{t-1}) \\ 0 & \rho \end{pmatrix} \begin{pmatrix} k_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} V(\phi_{t-1}) \\ 1 \end{pmatrix} \varepsilon_t. \quad (58)$$

For the PFL model, the sole state variable is $x_t = z_t$. The ALM of k_t is given by

$$k_t = T(\phi_{t-1})x_{t-1} + V(\phi_{t-1})\varepsilon_t \quad (59)$$

and the law of motion for z_t , which is exogenous, is simply

$$z_t = \rho z_{t-1} + \varepsilon_t. \quad (60)$$

Using the realized observation k_t , the agents now run the regression

$$k_t = x'_{t-1}\phi + \eta_t \quad (61)$$

to obtain a new estimate of ϕ_t . An appropriate estimation is ordinary least squares, which yields

$$\phi_t = \left(\sum_{i=1}^t x_{i-1}x'_{i-1} \right)^{-1} \sum_{i=1}^t x_{i-1}k_i. \quad (62)$$

To write the recursive expression of the OLS estimator, let

$$S_t = \sum_{i=1}^t x_{i-1}x'_{i-1}.$$

Then, we have that

$$S_t = S_{t-1} + x_{t-1}x'_{t-1}, \quad (63)$$

$$\phi_t = \phi_{t-1} + S_t^{-1}x_{t-1}(k_t - x'_{t-1}\phi_{t-1}). \quad (64)$$

To write the above recursive least squares algorithm as it usually appears in the adaptive learning literature, we can either define $R_t = S_t/t$ or $P_t = S_{t+1}/t$. Using the definition of

R_t , we get

$$\begin{aligned} R_t &\equiv \frac{1}{t}S_t = \frac{1}{t}(S_{t-1} + x_{t-1}x'_{t-1}) \\ &= \frac{1}{t-1}S_{t-1} - \frac{1}{t(t-1)}S_{t-1} + \frac{1}{t}x_{t-1}x'_{t-1} = R_{t-1} + \frac{1}{t}(x_{t-1}x'_{t-1} - R_{t-1}) \end{aligned} \quad (65)$$

and the recursion becomes

$$R_t = R_{t-1} + \frac{1}{t}(x_{t-1}x'_{t-1} - R_{t-1}), \quad (\text{RLS-1})$$

$$\phi_t = \phi_{t-1} + \frac{1}{t}R_t^{-1}x_{t-1}(k_t - x'_{t-1}\phi_{t-1}). \quad (\text{RLS-2})$$

Furthermore, using the definition of P_t , we get

$$P_t = \frac{1}{t}S_{t+1} = \frac{1}{t}(S_t + x_t x'_t) = \frac{1}{t-1}S_t - \frac{1}{t(t-1)}S_t + \frac{1}{t}x_t x'_t = P_{t-1} + \frac{1}{t}(x_t x'_t - P_{t-1}) \quad (66)$$

and the recursion becomes

$$\phi_t = \phi_{t-1} + \frac{1}{t}P_{t-1}^{-1}x_{t-1}(k_t - x'_{t-1}\phi_{t-1}), \quad (67)$$

$$P_t = P_{t-1} + \frac{1}{t}(x_t x'_t - P_{t-1}). \quad (68)$$

In the present paper, we will use the expressions (RLS-1) and (RLS-2) for the numerical implementation of the recursive least squares learning algorithm, but the two ways of writing the algorithm are essentially equivalent.¹²

5.1. Asymptotic Properties of RLS. The asymptotic properties of recursive least squares algorithms are well established in the literature. Here, we give a brief review of the main results concerning the convergence of the algorithm to the rational expectations solution and apply them to the reduced form of interest. First note that, using the actual law of motion, (RLS-1) and (RLS-2) can be rewritten as

$$R_t = R_{t-1} + \frac{1}{t}(x_{t-1}x'_{t-1} - R_{t-1}), \quad (69)$$

$$\phi_t = \phi_{t-1} + \frac{1}{t}R_t^{-1}x_{t-1}(x'_{t-1}(T(\phi_{t-1}) - \phi_{t-1}) + V(\phi_{t-1})\varepsilon_t). \quad (70)$$

For given initial conditions, it has been shown that the above algorithm converges locally to an REE of the reduced form (1)-(2) if and only if the REE is expectationally stable, or *E-stable*.¹³ By definition, a MSV solution of (1)-(2) is E-stable if it is a locally asymptotically

¹²The expressions (RLS-1) and (RLS-2) do not look precisely like (56), but they can be transformed into such an expression by stacking all the elements of ϕ and R into a column vector. This is important only for verifying that the ODE approach can be used, but is not necessary for the numerical implementation of least squares.

¹³The concept of E-stability is extensively discussed in Evans and Honkapohja (2001). Moreover, we postpone the discussion of the initial conditions of the algorithm, since it will be covered extensively in section 5.2.

stable equilibrium of the ordinary differential equation

$$\frac{d\phi}{d\tau} = T(\phi) - \phi. \quad (71)$$

Equivalently, an REE is E-stable if the Jacobian of $T(\phi) - \phi$ evaluated at the REE, given by

$$J^{LS}(\bar{\phi}) = \left. \frac{dT(\phi) - \phi}{d\phi} \right|_{\phi=\bar{\phi}} = \left. \frac{dT(\phi)}{d\phi} \right|_{\phi=\bar{\phi}} - I, \quad (72)$$

is a stable matrix, implying that it has eigenvalues with strictly negative real parts.

It has been shown by Marcet and Sargent (1989) that convergence of RLS learning can be achieved with *probability one* if the algorithm is augmented by a projection facility. This variation of the algorithm can be generally written as

$$\hat{R}_t = \hat{R}_{t-1} + \frac{1}{t} (x_{t-1}x'_{t-1} - \hat{R}_{t-1}), \quad (73)$$

$$\hat{\phi}_t = \hat{\phi}_{t-1} + \frac{1}{t} \hat{R}_t^{-1} x_{t-1} \left(x'_{t-1} (T(\hat{\phi}_{t-1}) - \hat{\phi}_{t-1}) + V(\hat{\phi}_{t-1}) \varepsilon_t \right), \quad (74)$$

$$\left(\hat{\phi}_t, \hat{R}_t \right) = \begin{cases} (\phi_t, R_t) & \text{if } (\phi_t, R_t) \in D_1 \\ \text{some value in } D_2 \subset D_1 & \text{if } (\phi_t, R_t) \notin D_1 \end{cases}. \quad (75)$$

Loosely speaking, the projection facility prevents the estimates from leaving a predetermined region D_1 . Thus, one can interpret this as a version of least squares learning that ignores observations that don't agree with the agents' priors. For instance, in some of our examples, it is reasonable to assume that agents should ignore observations that violate the stationarity condition $|\phi_k| < 1$. Also, if k can be interpreted as capital, then it may be reasonable to assume that agents know that $\phi_k > 0$. Typically, if these conditions are violated, i.e. if (ϕ_t, R_t) is not in D_1 , a natural choice for $(\hat{\phi}_t, \hat{R}_t)$ is the value of the previous period's estimates $(\hat{\phi}_{t-1}, \hat{R}_{t-1})$. It has been argued that the variation of algorithm with the projection facility is not always appropriate, as it may not always make economic sense. However, if the projection facility is not invoked, convergence with *probability one* cannot be shown. A general result regarding the convergence of (RLS-1) and (RLS-2) has been shown by Evans and Honkapohja (1998). The authors prove that, if $\bar{\phi}$ is E-stable, then the algorithm (RLS-1) and (RLS-2) converges to $\bar{\phi}$ with some probability which is bounded from below by a sequence of numbers tending to one.

Using the previous definition of E-stability, the conditions that need to be satisfied for the FLL reduced form (1)-(2) are:

$$\frac{a_1 a_2}{(1 - a_1 \bar{\phi}_1)^2} < 1 \quad \text{and} \quad \frac{\rho a_1}{1 - a_1 \bar{\phi}_1} < 1 \quad (76)$$

It can also be shown that, if the parameters of the model are such that $|a_1 + a_2| < 1$, the unique stationary REE solution is always E-stable, and the RLS algorithm always converges (see Evans and Honkapohja, 2001).¹⁴

¹⁴A similar result can be found in McCallum (2006) for multivariate models: it is shown that if current

For the PFL model, the E-stability condition is $a_1\rho < 1$. Note that for many well known models, parameter restrictions imply that this is always true (e.g. for the Lucas tree model, $a_1 = \beta \in (0, 1)$ and thus $a_1\rho$ is always less than one). Moreover, the REE of this reduced form model can be shown to be globally stable under learning (see Evans and Honkapohja (2001)). Therefore it is not necessary to employ the projection facility in order to ensure convergence.

5.2. Initial Conditions for RLS. Next, we describe how to initialize the recursive least square algorithm. Although, theoretically, the effect of the initial conditions disappears in the limit, the initial conditions for ϕ and R may be crucial if one is interested in short and medium run dynamics. It should be emphasized that this is particularly relevant for macroeconomic modeling, since macroeconomic time series data, unlike data series used in other fields like engineering or finance, typically have short length.¹⁵

The importance of the initial values will also become clearer in section 8, where we compare realizations of the RLS algorithm for a variety of initial conditions within the context of the examples introduced earlier. In what follows, we discuss three different ways of setting the initial values. The first derives in a natural way from an estimation based on randomly generated data before the initial time period of learning. The second is a more ad-hoc way of choosing the initial values. The third is a variant of the first, where the initial conditions are drawn from an appropriate distribution.

Before proceeding with the three ways of initializing the RLS algorithm, we should point out that there are two issues that one should be cautious about when setting up the recursion. First, when deriving the recursive formula for the OLS, in order to be able to do the steps in (65), we must have that $t \neq 1$.¹⁶ Second, for the regression (61) to make sense and the OLS estimate (62) to be well-defined, the matrix S_t , or in effect the matrices R_t or P_t , must be invertible. A necessary condition for invertibility of S_t is that the sample size t is larger than or equal to the number of regressors. For the regression equation corresponding to the FLL reduced form, the two regressors are k_{t-1} and z_{t-1} , whereas the only regressor for the PFL reduced form is z_{t-1} . Given these considerations, if t_0 denotes the time period for which we set the initial values of the recursive algorithm, we clearly must have that $t_0 \geq 2$. In other words, *the recursive formulation of the OLS (RLS-1, RLS-2) for the model at hand is only valid for $t \geq 2$.*

values of endogenous variables are included in individuals' information sets, then determinacy (i.e. unique stationary MSV solution) implies E-stability.

¹⁵The issue of initializing learning algorithms is also addressed in the work of Milani (2005a, 2005b, 2005c), Orphanides and Williams (2005a, 2005b) and Williams (2003). Furthermore, Evans and Honkapohja (2001, section 2.6) provide a brief discussion on the initial conditions of recursive least squares.

¹⁶Apart from the fact that we cannot proceed with the steps in expression (65) for $t = 1$, the recursion would not give an invertible matrix R for all possible initial values of the state variable. For example, if $x_0 = (0, 0)$, then according to the recursive formula $R_1 = 0$.

Initial Conditions from Randomly Generated Data (RGD). One way of defining initial conditions that is compatible with the recursive algorithm is as follows

$$S_{t_0} = \sum_{i=1}^{t_0} x_{i-1}x'_{i-1} \text{ and } \phi_{t_0} = S_{t_0}^{-1} \sum_{i=1}^{t_0} x_{i-1}k_i, \quad (77)$$

or

$$R_{t_0} = \frac{1}{t_0} \sum_{i=1}^{t_0} x_{i-1}x'_{i-1} \text{ and } \phi_{t_0} = \frac{1}{t_0} R_{t_0}^{-1} \sum_{i=1}^{t_0} x_{i-1}k_i. \quad (78)$$

Two interrelated issues arise here. First, since the recursive algorithm starts at $t_0 + 1$, we should determine how the observations x_i , for $i \in \{1, \dots, t_0\}$ are generated given the initial x_0 . Second, while t_0 has to be larger than or equal to the number of regressors, it is not clear how large it should be. Note that, once the first issue is resolved and some initial data are generated, it is easy to work out the first time period that renders S_t invertible and then use the corresponding time period as t_0 .

Regarding how the initial data are generated, there is no problem when the model at hand is the PFL reduced form. Since the state variable is exogenous and known, we can simply generate t_0 observations and calculate the corresponding OLS estimates. However, for the FLL reduced form, the way to generate the data is not obvious. One way to do this is by using the REE, i.e. given some initial conditions $(k_0, z_0)'$ for the state variables, we can obtain

$$x_t = \begin{pmatrix} k_t \\ z_t \end{pmatrix} = \begin{pmatrix} \bar{\phi}_k & \bar{\phi}_z \\ 0 & \rho \end{pmatrix} \begin{pmatrix} k_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} \bar{\phi}_z/\rho \\ 1 \end{pmatrix} \varepsilon_t, \quad (79)$$

for all $t \in I_0 = \{1, \dots, t_0\}$ and then use these values to get R_{t_0} and ϕ_{t_0} from (78). We will return to the issue of generating the initial data later in this section.

With these initial conditions, the recursive algorithm becomes

$$R_{t_0} = \frac{1}{t_0} \sum_{i=1}^{t_0} x_{i-1}(\bar{\phi})x'_{i-1}(\bar{\phi}), \quad (80)$$

$$\phi_{t_0} = \frac{1}{t_0} R_{t_0}^{-1} \sum_{i=1}^{t_0} x_{i-1}(\bar{\phi})k_i(\bar{\phi}), \quad (81)$$

$$\left\{ \begin{array}{l} R_t = R_{t-1} + \frac{1}{t} (x_{t-1}x'_{t-1} - R_{t-1}) \\ \phi_t = \phi_{t-1} + \frac{1}{t} R_t^{-1} x_{t-1} (k_t - x'_{t-1}\phi_{t-1}) \end{array} \text{ for } t \in I_1 = \{t_0 + 1, t_0 + 2, \dots\} \right\}, \quad (82)$$

where the notation $x_i(\bar{\phi})$ is used to emphasize the fact that the regressors are generated using the REE.

Ad-hoc Initial Conditions (AH). It is quite common to initialize the recursive equations of the algorithm at $t_0 = 0$ for some arbitrarily chosen invertible matrix S_0 and some ϕ_0 . Note that for $t_0 = 0$ and arbitrary initial conditions S_0 and ϕ_0 , the recursive estimate at

time t is equal to

$$\phi_t = \left(S_0 + \sum_{i=1}^t x_{i-1} x'_{i-1} \right)^{-1} \left(S_0 \phi_0 + \sum_{i=1}^t x_{i-1} k_i \right). \quad (83)$$

This expression can be verified easily by using the fact that $\phi_t = S_t^{-1}(S_{t-1}\phi_{t-1} + x_{t-1}k_t)$ and first substituting $S_{t-1}\phi_{t-1}$ backwards until $S_0\phi_0$ is reached and then substituting for S_t^{-1} using the fact that $S_t = S_{t-1} + x_{t-1}x'_{t-1} = \dots = S_0 + \sum_{i=1}^t x_{i-1}x'_{i-1}$. This implies that, for the recursive estimate to be close to the off-line estimate, S_0 would have to be a relatively "small" matrix.¹⁷ Of course, as t becomes very large, the effect of the initial conditions should theoretically disappear.

With the ad-hoc initial conditions, the recursive algorithm becomes

$$\begin{cases} R_1 = \frac{1}{1}S_1 = S_0 + x_0x'_0 \\ \phi_1 = \phi_0 + R_1^{-1}x_0(k_1 - x'_0\phi_0) \end{cases}, \quad (84)$$

$$\begin{cases} R_t = R_{t-1} + \frac{1}{t}(x_{t-1}x'_{t-1} - R_{t-1}) \\ \phi_t = \phi_{t-1} + \frac{1}{t}R_t^{-1}x_{t-1}(k_t - x'_{t-1}\phi_{t-1}) \end{cases} \quad \text{for } t \in \{2, 3, \dots\}, \quad (85)$$

$$S_0 \text{ and } \phi_0 \text{ given.} \quad (86)$$

Initial conditions drawn from a distribution (DIS). Our last method of initializing the algorithm is a variant of the RGD method. The basic idea is that the initial condition for ϕ is drawn from an appropriate distribution, centered around the REE, $\bar{\phi}$. As is well known, the OLS estimator is normally distributed with a mean equal to the true parameter (here $\bar{\phi}$) and a variance that depends on the regressors and the variance of the disturbance terms. Therefore a suitable distribution will be a normal with mean $\bar{\phi}$. Regarding the variance, we propose using the asymptotic variance, which can be interpreted as the variance of the asymptotic approximation of the finite sample distribution (i.e. the variance expected over repeated sampling).¹⁸

To determine the asymptotic variance, note first that, from the law of large numbers,

$$p \lim_{t_0 \rightarrow \infty} R_{t_0} = p \lim_{t_0 \rightarrow \infty} \left(\frac{1}{t_0} \sum_{i=1}^{t_0} x_{i-1}(\bar{\phi}) x'_{i-1}(\bar{\phi}) \right) = E(x_{i-1}(\bar{\phi}) x'_{i-1}(\bar{\phi})) = M(\bar{\phi}). \quad (87)$$

¹⁷The terms *on-line* and *off-line* estimation are borrowed from the engineering literature. Off-line refers to estimation in the usual sense, i.e. the econometrician collects the data after, say, T periods and these yield an estimate using expression (62). On-line estimation is simply recursive estimation: in any given period up to T , the econometrician updates the estimate using the new observation.

¹⁸It is worth emphasizing that these normality results hold true only for constant underlying parameters to be estimated, i.e. for rational expectations dynamics. Such results do not hold in general along learning paths.

For the FLL reduced form, $M(\bar{\phi})$ is given by¹⁹

$$M(\bar{\phi}) = \frac{\sigma^2}{1 - \rho^2} \begin{pmatrix} \frac{b^2(1 + \rho\bar{\phi}_k)}{(1 - a_1\rho - a_1\bar{\phi}_k)^2(1 - \rho\bar{\phi}_k)(1 - \bar{\phi}_k^2)} & \frac{b}{(1 - a_1\rho - a_1\bar{\phi}_k)(1 - \rho\bar{\phi}_k)} \\ \frac{b}{(1 - a_1\rho - a_1\bar{\phi}_k)(1 - \rho\bar{\phi}_k)} & 1 \end{pmatrix}. \quad (88)$$

For the PFL model, $M(\bar{\phi})$ is equal to

$$M(\bar{\phi}) = \frac{\sigma^2}{1 - \rho^2}. \quad (89)$$

The asymptotic distribution of the OLS estimate is then given by

$$\phi_{t_0} \stackrel{a}{\sim} N\left(\bar{\phi}, \frac{\sigma_\eta^2}{t_0} M^{-1}(\bar{\phi})\right), \quad (90)$$

where

$$\sigma_\eta^2 = \text{Var}(\eta_t) = \text{Var}(V(\bar{\phi})\varepsilon_t) = V(\bar{\phi})^2\sigma^2. \quad (91)$$

Note that $\sigma_\eta^2 = \frac{\bar{\phi}_z^2\sigma^2}{\rho^2}$ for the FLL reduced form and $\sigma_\eta^2 = \frac{\bar{\phi}^2\sigma^2}{\rho^2}$ for the PFL reduced form. For a given t_0 , one can then generate a random realization of ϕ_{t_0} using the above distribution and then use this as an initial value. Furthermore, we can reset the set of indices I_1 to $I'_1 = \{1, 2, \dots\}$ and rewrite the recursive algorithm (80)-(82) as

$$R_0 = R_{t_0} = M(\bar{\phi}), \quad (92)$$

$$\phi_0 = \phi_{t_0} : \text{generated randomly from } N\left(\bar{\phi}, \frac{\sigma_\eta^2}{t_0} M^{-1}(\bar{\phi})\right), \quad (93)$$

$$\left\{ \begin{array}{l} R_t = R_{t-1} + \frac{1}{t+t_0} (x_{t-1}x'_{t-1} - R_{t-1}) \\ \phi_t = \phi_{t-1} + \frac{1}{t+t_0} R_t^{-1} x_{t-1} (k_t - x'_{t-1}\phi_{t-1}) \end{array} \text{ for } t \in I'_1 = \{1, 2, \dots\} \right\}. \quad (94)$$

In general, this way of producing initial conditions produces a less extreme initial ϕ than the RGD method, since the asymptotic variance will typically be smaller than the variance that results from one realization of the regressors.

Comments. The asymptotic properties of the recursive least squares algorithm hold locally regardless of the way the initial conditions are chosen. In practical terms, this means that, if the initial conditions are very far from the limiting point of the algorithm (or the learnable REE), there is a positive probability that the estimates ϕ_t might not converge to $\bar{\phi}$. In this sense, choosing AH or DIS initial conditions might be better than RGD, if one wants to increase the chance of convergence to the REE. However, the variability of the recursive estimates can be considerably reduced if one uses the RGD as long as t_0 is large. To see this,

¹⁹The asymptotic moment matrix $M(\bar{\phi})$ is well-defined, since the solution for the endogenous variable k_t is stationary by assumption. For a derivation of the matrix see Giannitsarou (2005).

note that the variance of ϕ_{t_0} and the updating step

$$\frac{1}{t + t_0} R_t^{-1} x_{t-1} (k_t - x'_{t-1} \phi_{t-1}) \quad (95)$$

in the recursive algorithm expression become smaller as t_0 increases, reducing the chance of the estimate deviating very much from the rational expectations equilibrium. In addition, a large t_0 for the RGD would give a similar initial condition to the one produced by DIS, since the latter is the asymptotic approximation of the former. On the other hand, if one sets the initial conditions by generating random data with a small t_0 , it is possible that the initial estimate ϕ_{t_0} is imprecise and too far away from $\bar{\phi}$ (due to the small sample size), thus leading to divergence. In particular, for the FLL reduced form, and if the rational expectations solution $\bar{\phi}_k$ is very close to one (near unit root), the initial period estimate with RGD might well be above one, in which case the subsequent estimates might turn out to be explosive. One way to overcome this problem is to ensure that ϕ_{k,t_0} satisfies at least the stationarity condition $|\phi_{k,t_0}| < 1$. However, imposing the stationarity condition may lead to a certain type of "bias" in the estimates ϕ and consequently in the variables of the model, an issue which we will discuss in more detail in section 8, where the numerical results are presented.

Second, given the self-referential nature of the models that we study, choosing the initial conditions based on observations generated by the REE (RGD) creates some conceptual problems for the reduced form model with the lag, since in order for the data to be generated with the REE, the agents have to form correct expectations about the endogenous variable before they start learning about it. However, this is not an issue for the PFL model since it contains only variables that are determined exogenously, and doing an on-line estimation is equivalent to doing an off-line estimation once all data are collected. In this sense, one could argue that the ad-hoc specification is preferable for the FLL reduced form. On the other hand, it would still be difficult to decide which initial values are appropriate, and choosing $\phi_0 = \bar{\phi}$ and $R_0 = M(\bar{\phi})$ is implicitly pre-supposing that agents have estimated and found the REE before starting to learn about it.

Despite these considerations, there are good arguments in favor of all methods for initializing, depending on the problem or situation the researcher wishes to study. Regarding the RGD method, we can argue that it is reasonable to assume that it takes some time before collected data may become useful for statistical analysis. In fact, the development and evolution of econometrics as a part of economics was mainly driven by the increased availability of data in the early years of the twentieth century. In this respect, the interpretation we are suggesting is that, before starting to forecast, agents collect data from macroeconomic time series until these become useful for estimation, in the sense that t_0 becomes large enough to give well-behaved econometric estimates. The AH method could be useful when studying structural changes. Suppose, for instance, that there is a structural change at time t_0 which shifts one of the parameters of the model, resulting in a change of the rational expectations equilibrium and possibly of the steady state of the economy. If such a change is not announced, one can then suppose that, up to time t_0 , agents have a series of data based on

the old equilibrium, which could be long enough to yield consistent estimates. Since they cannot observe that there is a structural change, it is then reasonable to assume that the agents continue to forecast the variables at time t_0 using these series, and thus an initial ϕ that corresponds to the old equilibrium.²⁰ Furthermore, the AH method is useful in an applied context, when the initial conditions can be estimated from data.²¹ Finally, regarding the DIS method, it can be argued that it is the suitable method for initializing in a context where there is heterogeneity in initial priors in the economy. The DIS method implies that the initial value is $\bar{\phi}$ on average, but allows for the possibility that at a given point in time (where learning starts) not all the population necessarily has the same belief about the coefficient ϕ (reflecting for example pessimism or optimism).

6. STOCHASTIC GRADIENT (SG)

The second algorithm of interest is the so-called stochastic gradient or least mean squares learning. This algorithm has been studied in the literature as a simple alternative to least squares. Although it is an inefficient estimator, the advantage of SG learning is its computational simplicity, especially in multivariate and complicated models.²² The stochastic gradient recursive algorithm is given by

$$\phi_t = \phi_{t-1} + \frac{1}{t} x_{t-1} [x'_{t-1} (T(\phi_{t-1}) - \phi_{t-1}) + V(\phi_{t-1})\varepsilon_t]. \quad (\text{SG})$$

The algorithm differs from least squares in that it ignores the second moment matrix when updating.

6.1. Asymptotic Properties of SG. It can be shown that an MSV solution of the reduced form models is learnable under stochastic gradient if it is a locally asymptotically stable equilibrium of the ordinary differential equation

$$\frac{d\phi}{d\tau} = [M(\phi) (T(\phi) - \phi)], \quad (96)$$

where $M(\phi) = \lim_{t \rightarrow \infty} E [x_t(\phi)x_t(\phi)']$. The local asymptotic stability of an REE $\bar{\phi}$ under stochastic gradient learning is determined by the stability of the Jacobian matrix $J^{SG}(\phi) = d[M(\phi) (T(\phi) - \phi)]/d\phi$, evaluated at $\bar{\phi}$. The stochastic gradient algorithm converges locally to the REE if and only if the real parts of the eigenvalues of $J^{SG}(\bar{\phi})$ are strictly negative. For the FLL reduced form, it is not possible to present these conditions in an elegant manner, but it can be shown that E-stability ensures convergence of the SG algorithm, whenever there exists a *unique* stationary MSV solution, i.e. whenever $|a_1 + a_2| < 1$. For the (univariate) PFL reduced form, it can easily be shown that the condition for convergence of the stochastic gradient algorithm is given by the E-stability condition. However, if the reduced form is

²⁰As an example, Giannitsarou (2006) studies the short-run learning dynamics of a tax reform in the context of example two, assuming that the tax on capital is reduced to zero at t_0 . The initial ϕ is set below $\bar{\phi}$ and corresponds to the pre-reform equilibrium.

²¹For example, Orphanides and Williams (2005a) estimate the initial conditions from data that range from 1948 to 1965, and then simulate their learning model implicitly assuming that learning starts in 1965.

²²For a discussion and references on SG learning, see Evans, Honkapohja and Williams (2006).

multivariate, then E-stability does not necessarily imply convergence of the SG algorithm (see Evans, Honkapohja and Williams, 2006). Once again, convergence with probability one is attained only by augmenting the algorithm with a projection facility as in section 5.1.

6.2. Initial Conditions for SG. Unlike the RLS, there are no technical complications related to the invertibility of the second moment matrix when setting the initial conditions for the stochastic gradient algorithm. Here, the recursion (SG) is well defined for any $t = 1, 2, \dots$, given some initial condition ϕ_0 . The algorithm will be initialized with the AH method.

7. CONSTANT GAIN (CG)

Popular alternatives to the standard RLS and SG algorithms are their constant gain or tracking variants, i.e. algorithms which replace the decreasing gain sequence $1/t$ with a constant positive and small number $0 < g < 1$.²³ In particular these are

$$\begin{aligned} R_t &= R_{t-1} + g(x_{t-1}x'_{t-1} - R_{t-1}), \\ \phi_t &= \phi_{t-1} + gR_t^{-1}x_{t-1}(k_t - x'_{t-1}\phi_{t-1}) \end{aligned} \tag{CGRLS}$$

and

$$\phi_t = \phi_{t-1} + gx_{t-1}(k_t - x'_{t-1}\phi_{t-1}). \tag{CGSG}$$

The constant gain recursive least squares algorithm (CG-RLS) can be derived from its non-recursive counterpart, in a similar way as the regular RLS. To see this, we start from the non-recursive constant gain least squares estimator

$$\phi_t = \frac{1}{g} \left[\sum_{i=1}^t (1-g)^{i-1} x_{t-i} x'_{t-i} \right]^{-1} \left[\sum_{i=1}^t (1-g)^{i-1} x_{t-i} k_{t-i+1} \right]. \tag{97}$$

Let

$$R_t = g \sum_{i=1}^t (1-g)^{i-1} x_{t-i} x'_{t-i}, \tag{98}$$

$$A_t = \sum_{i=1}^t (1-g)^{i-1} x_{t-i} k_{t-i+1}. \tag{99}$$

First, it is easy to verify that

$$R_t = (1-g)R_{t-1} + gx_{t-1}x'_{t-1}, \tag{100}$$

$$A_t = (1-g)A_{t-1} + x_{t-1}k_t, \tag{101}$$

so that

$$A_{t-1} = \frac{1}{1-g} (A_t - x_{t-1}k_t). \tag{102}$$

²³Constant gain algorithms have been recently adopted in empirical papers such as Bullard and Eusepi (2005), Milani (2005a, 2005b, 2005c), and Orphanides and Williams (2005a, 2005b), among others.

Next, we have that

$$R_{t-1}\phi_{t-1} = A_{t-1} = \frac{1}{1-g} (A_t - x_{t-1}k_t). \quad (103)$$

Finally, to find the recursive constant gains least squares algorithm, we use expressions (100)-(103) to get

$$\begin{aligned} \phi_t &= R_t^{-1}A_t \\ &= R_t^{-1}[(1-g)A_{t-1} + x_{t-1}k_t] \\ &= R_t^{-1}[(1-g)R_{t-1}\phi_{t-1} + x_{t-1}k_t] \\ &= R_t^{-1}[(R_t - gx_{t-1}x'_{t-1})\phi_{t-1} + x_{t-1}k_t] \\ &= \phi_{t-1} - gR_t^{-1}x_{t-1}x'_{t-1}\phi_{t-1} + R_t^{-1}x_{t-1}k_t \\ &= \phi_{t-1} + gR_t^{-1}x_{t-1}(k_t - x'_{t-1}\phi_{t-1}). \end{aligned} \quad (104)$$

Summing up, the recursive algorithm is

$$R_t = R_{t-1} + g(x_{t-1}x'_{t-1} - R_{t-1}), \quad (105)$$

$$\phi_t = \phi_{t-1} + gR_t^{-1}x_{t-1}(k_t - x'_{t-1}\phi_{t-1}). \quad (106)$$

The main practical difference between these algorithms and the regular RLS and SG is that they do not allow for convergence to a constant as $t \rightarrow \infty$. This is clear from the expressions (CG-RLS) and (CG-SG) where it can be seen that the forecasting error is always added to the updated estimate, even in the limit, while in the regular algorithms the effect of the current forecasting error vanishes in the limit. Another way to see the difference between the two algorithms is by considering their non-recursive expressions. Regular RLS (or SG) puts equal weights on all observations across time; as $t \rightarrow \infty$ the weight assigned to each observation becomes very small and therefore the effect of an outlier to the estimate is marginal, allowing the estimate to converge to the true parameter value $\bar{\phi}$. On the other hand, the CG-RLS algorithm assigns decreasing weights on observations across time, so that recent observations matter for the current estimate even in the limit.

For these reasons, it is a common belief that constant gain algorithms are particularly useful in tracking structural changes in the model and have been consequently used widely in models with policy making where the policy is allowed to shift over time.²⁴ Constant gain algorithms are also relevant when the modeling assumptions require that the forecasters prefer to use more recent data. Such an example is stock price forecasting (see e.g. Carceles-Poveda and Giannitsarou, 2006).

7.1. Asymptotic properties of CG. The asymptotic properties of constant gain learning differ substantially from the ones of decreasing gain learning, since under CG there is a non-zero correction for the estimated coefficients through the forecasting error, even in the limit as $t \rightarrow \infty$. This means that asymptotically, the constant gain algorithm converges to

²⁴Marcet and Nicolini (2003) provide a good discussion on the virtues and disadvantages of decreasing and tracking algorithms depending on the modeling context.

a distribution rather than the constant $\bar{\phi}$. Indeed, Evans and Honkapohja (2001) show that under certain conditions and assumptions, a constant gain algorithm converges in distribution to a normal centered around the REE, $\bar{\phi}$. Although it is not possible to derive clean conditions that ensure convergence, such as the E-stability conditions for RLS, it is known that two necessary conditions for convergence of CG algorithms in distribution are (a) that $0 < g < 1$ and (b) the E-stability conditions.²⁵

7.2. Initial conditions for CG. As with RLS, when deriving the recursive expression for the constant gain (least squares) algorithm we implicitly made some assumptions that need to be satisfied in order for the estimator to be meaningful. In principle, the same considerations regarding the initial conditions arise here too, i.e. that t_0 must be at least 2 and such that the matrix R is invertible. Furthermore, the gain needs to be strictly between 0 and 1 in order to imply finite sums in the limit $t \rightarrow \infty$. The values 0 and 1 for the gain must be excluded, since otherwise, from (97), we would have that $\phi_t = 0$. Once again, we allow for the same three possible ways of initializing, as with RLS, namely RGD, AH or DIS.

Nevertheless, the choice of initial conditions for the constant gain algorithm will prove less critical than for the RLS, a fact that will become more apparent once we present the numerical results. As long as the initial conditions are not very far from the REE and the gain is not very large, the estimates will fluctuate around the REE with a stationary distribution. In contrast, with RLS, the choice of initial conditions is very important, since the updating is rather slow and the initial condition will determine from which direction the estimate approaches the REE (this is critical for short horizons). Because of these reasons, and because a short sample estimator based on (97) might have very high variance, yielding misleading results, we prefer to generate the initial conditions based on simple OLS estimates for the RGD and DIS methods.

8. NUMERICAL RESULTS

This section presents some numerical results based on the two examples (stochastic growth model and Lucas tree model) using the learning algorithms discussed in the previous sections. In particular, we explore the sensitivity of the dynamics of the two models to different learning algorithms, as well as to different ways of choosing the initial values of ϕ . For both examples, we present two types of numerical results: (a) some figures of the evolution of coefficient estimates and key variables of the models plotted together with the corresponding evolution of the variables under rational expectations and (b) statistics for the variables based on 500 replications of each experiment.

For the first example, since the reduced form comes from the log-linearization of a non-linear model, we restrict estimates and the initial values in all the experiments to be stationary. That is, we require that the implied ϕ_{t_0} is less than one in absolute value and we include a projection facility that only allows for stationary coefficients along the recursion. Moreover, whenever we initialize the RLS algorithms with RGD, in order to find the minimum t_0 , we require that S_{t_0} is invertible. We use the standard quarterly parametrization

²⁵ Evans and Honkapohja (2001) provide a set of more elaborate conditions that ensure convergence.

for the stochastic growth model, i.e. $\alpha = 0.36$, $\beta = 0.99$, $\delta = 0.025$, $\gamma = 1$, $\rho = 0.95$ and $\sigma = 0.00712$. For this parameterization, the rational expectations law of motion for capital (REE) is

$$k_t = 0.9653k_{t-1} + 0.0716z_{t-1} + 0.0756\varepsilon_t, \quad (107)$$

i.e. $\bar{\phi}_k = 0.9653$ and $\bar{\phi}_z = 0.0716$. Note that, for this example, the persistence coefficient $\bar{\phi}_k$ is close to one, a fact that will have important implications for the results we obtain.

For the second example, i.e. the Lucas tree model, we assume that $\beta = 0.99$ and $\gamma = 1$. In addition, the two parameters of the dividend process are set to $\rho = 0.95$ and $\sigma = 0.06$, corresponding to the slope coefficient and error variance obtained when regressing the logged, seasonally adjusted quarterly dividends series in the data on its first lag. The rational expectations solution in this case is $\bar{\phi} = 0.95$. As stated earlier, no projection facility is necessary when computing the dynamics of the Lucas tree model.

Finally, for both examples, the disturbance terms for the exogenous state variable are generated from a normal distribution. In order to make the results comparable, we have used the same state for the normal distribution generator of Matlab, which was set to 56. All the experiments are done for a time horizon of $T = 150$ periods, representing quarters. Finally, the statistics from all experiments are generated by applying the Hodrick-Prescott filter for quarterly data.

8.1. The Stochastic Growth Model.

RLS: Results based on RGD, AH and DIS initializations. We start by discussing the results under RLS based on RGD and DIS initializations. To determine the minimum t_0 for generating the data during the training period of the RGD initialization, we require that the program looks for the first t_0 such both that the determinant of S_{t_0} is larger than the default tolerance level of Matlab (invertibility condition) and that the corresponding estimate ϕ_{t_0} that multiplies the lag is less than one in absolute value (stationarity condition). Furthermore, the initial estimate of the DIS initialization is drawn from a distribution that approximates the variance of the OLS estimator of ϕ based on 5 observations.

Figure 2 shows the evolution of capital for four different specifications of the initial value, namely for t_0 being the minimum required for the matrix S_{t_0} to be invertible (for the particular realization of the shock this turns out to be $t_0 = 6$), and for $t_0 = 15, 40, 100$ periods. As expected, the larger t_0 is, the closer the adaptive learning dynamics are to those of rational expectations. This is because, as t_0 increases, there are two effects which contribute to the evolution of adaptive learning. First, the larger the t_0 , the better is the initial value estimate of ϕ , in the sense that it is closer to the REE value $\bar{\phi}$. Second, the larger the t_0 , the smaller is the gain sequence $\frac{1}{t+t_0}$; in other words, the weight put on the forecasting error is smaller and it is therefore more difficult for the estimate to move far away from the REE. It is worth noting that, when t_0 is the smallest possible, the log deviations of the variables from the steady state with RE and learning can differ up to 2 percentage points for capital and consumption, and up to 7.5 percentage points for investment.

< FIGURE 2 ABOUT HERE >

A second observation is that capital (and therefore investment) appears to be less volatile under learning than under rational expectations. Consequently, consumption will be more volatile. This effect is bigger the smaller t_0 is. Since the figures show the dynamics for only one realization of the shock, we have calculated statistics for the variables based on many (500) replications of the experiment.

< TABLE 1 ABOUT HERE >

Table 1 presents the average relative deviations and correlations with output for capital, investment and consumption. The first four rows present the statistics for the RGD initialization, row five presents the results for the DIS initialization and the last row displays the statistics under RE. A few comments on the entries of rows 1-4: First, as already illustrated by the figures, the smaller the t_0 , the less volatile capital and investment are under adaptive learning. As a consequence, consumption is more volatile relative to output. Second, the correlations with output under learning are also further away from their RE values with a smaller t_0 . In particular, capital has a higher correlation the smaller the t_0 is, and this difference is quite large with the minimum possible t_0 , whereas the corresponding correlations for investment and consumption are not so different. Third, while there is a significant difference between the statistics when t_0 is the minimum possible and the statistics under RE, the numbers suggest that these differences appear to be eliminated quite quickly as t_0 becomes larger. As we see, the statistics with $t_0 = 15$ are already quite close to those under RE. Finally, the numbers with the DIS initialization are relatively close to their RE counterparts. This is not surprising, since the DIS method should give similar results to the RGD initialization with a relatively large t_0 .

We have also performed the same experiments with the stricter projection facility that imposes non-negativity for ϕ_k and ϕ_z and we find that this does not affect the results in any significant way. Finally, we simulated the paths of the variables under learning for $T = 2000$, using RGD with t_0 being the minimum possible, and we find that the statistics are not very different from those that we get with $T = 150$. In particular, the relative deviation and correlation with output of the capital stock are equal to 0.1428 and 0.4375 respectively.

We now present the results based on AH initializations. Here, we focus on the statistics of capital under learning for a variety of different initial conditions that can be above, below or at the REE. These are given in Table 2, displaying the relative deviations and the corresponding correlations in parentheses. Note that the rational expectations statistics are 0.2619 (relative deviation) and 0.3604 (correlation with output).

< TABLE 2 ABOUT HERE >

From the table, it can be seen that the levels of the initial values for ϕ_k and ϕ_z do matter for the evolution of capital. First, moving along the diagonal, we see that increasing the values of the initial coefficients leads to a higher relative variability and a lower correlation with output, whereas starting from $\bar{\phi}$ generates statistics that are very close to their RE counterparts. Second, we see that ϕ_k matters more, especially for the variability of capital.

In particular, if $\phi_{k0} > \bar{\phi}_k$ then capital is more volatile under learning than under rational expectations. This is reflected in columns 1-3 of the table. In the first column, where $\phi_{k0} = 0.9\bar{\phi}_k$, the variability of capital is always below the RE value, even when the initial value of ϕ_{z0} is above $\bar{\phi}_z$. Similarly, along the the third column, where $\phi_{k0} = 1.025\bar{\phi}_k$, the variability of capital is always higher than the RE value, regardless of the value of ϕ_{z0} . Finally, whereas different initial values for ϕ_z also yield some differences, these are small compared to the effect of different initial values of ϕ_k . We will discuss these findings in more detail later in this section.²⁶

SG and CG: Comparison with RLS. Next, we do some experiments with SG learning, based on ad-hoc initial conditions, and with CG learning, using the same initializations as in previous subsections. The results for the SG algorithm with ad-hoc initializations are summarized in Table 3. Comparing the numbers for RLS and SG, it can be seen that, qualitatively, the results are not very different. As with RLS, the initial conditions may matter for the statistics of the aggregate variables. In particular, if the initial conditions are below the REE (and in particular the coefficient ϕ_k), the relative deviation of capital is lower, and the correlation is higher, and vice versa.

< TABLES 3 - 5 ABOUT HERE >

The results for the ad-hoc initialization under CG are displayed in Tables 4-5 for gain values of 0.02 and 0.2.²⁷ The tables reveal the following facts. First, the statistics with a gain of $g = 0.02$ are very similar to the ones obtained under RLS and SG. Again, we observe that higher values of the initial coefficient ϕ_{k0} seem to matter considerably for the relative variability and correlation with output of capital. Looking at the results with the higher gain of $g = 0.2$, however, we observe that the statistics are somewhat less affected by the initial coefficient values. In particular, the correlations with output are very similar across all initial values. In addition, the differences in the relative variabilities are smaller across initial values of ϕ_{k0} and ϕ_{z0} , while all the values are higher with the higher gain. As an example, the relative variability of capital starting from $\bar{\phi}$ is equal to 0.327 with a gain of 0.2, whereas it is equal to 0.2706 and 0.278 with a gain of 0.02 and RLS respectively. The statistics for the CG algorithm with the RGD and DIS initializations are displayed in Tables 6-7. The results with a gain of 0.02 and RGD or DIS initializations are again very similar to the RLS findings. In particular, we again observe that a higher gain coefficient leads to higher variability of capital in all cases considered.

< TABLES 6 - 7 ABOUT HERE >

²⁶ As with the RGD initialization, we have simulated the paths of the variables under learning for $T = 2000$, for the same initial conditions as in the short horizon experiments. Once again, we obtain very similar results, indicating that the effect of the initial conditions takes a long time to disappear.

²⁷ The choice of the gain parameter depends on the problem at hand and is an issue of applied nature. Various authors calibrate or estimate the gain parameter (e.g. see Milani, 2005a, 2005b, 2005c, Orphanides and Williams, 2005a, 2005b, and Carceles-Poveda and Giannitsarou, 2006).

Discussion. Collecting the results from the numerical experiments, one can confidently conclude that the initial values matter for the dynamics of aggregate macroeconomic variables in the stochastic growth model under adaptive learning, whereas the different algorithms do not seem to matter that much. A marked pattern from the experiments is that if the initial condition for ϕ_k is below the REE, then the volatility of the state variable under learning is lower than the volatility under rational expectations. The reason for this significant difference between the statistics relates to how these depend on the estimated elasticities ϕ_t . Under rational expectations, the elasticities are constant and it is straightforward to establish the monotonicity properties of the statistics with respect to $\bar{\phi}$. For example, it can be shown that the variance of the endogenous state variable under rational expectations is²⁸

$$E(k_t^2) \equiv \sigma_k^2 = \frac{\gamma^2(1 + \rho\bar{\phi}_k)}{(1 - a_1\rho - a_1\bar{\phi}_k)^2(1 - \rho\bar{\phi}_k)(1 - \bar{\phi}_k^2)} \quad (108)$$

and it is possible to show that this is increasing in $\bar{\phi}_k$, for all reasonably calibrated parameter values. Using this result, we can then heuristically argue that the variance of the endogenous state variable under learning will be lower than the variance under rational expectations if the estimated coefficients ϕ_{kt} remain well below $\bar{\phi}_k$ for a large number of periods. For example, Figure 3 shows the paths of ϕ_t that generate the variables in Figure ?? with the minimum t_0 , where it can be seen that the initial value of ϕ_k is below the REE and that the elasticity falls quite low and remains quite far from the REE for many periods before it starts converging. Furthermore, for every experiment that produces the statistics in Tables 1 - 3, we record the average percentage number of periods for which ϕ_k is below the REE. These percentages are displayed in Table 8 along with the corresponding relative deviation of capital under adaptive learning.

< FIGURE 3 ABOUT HERE >

< TABLE 8 ABOUT HERE >

The table reflects an important difference between the RLS and SG algorithms: because the SG algorithm does not update the second moment matrix, it will generate a very high persistence of the coefficient estimates. In turn, these remain above or below the REE (depending on the initial value) much more often than with RLS.

Some more comments are also in place regarding the pattern of the statistics when using the RGD initialization. For the particular model at hand, the REE is $\bar{\phi}_k = 0.9653$ and $\bar{\phi}_z = 0.0716$. In other words, the REE elasticity ϕ_k is very close to 1, which is the bound for ensuring stationarity. When generating the pre-learning data that is used to calculate the initial value for ϕ_k , we are looking for the minimum possible t_0 that satisfies both the invertibility and the stationarity conditions. Although this data is generated by using the REE, by not accepting a coefficient larger than one as an initial condition, we are implicitly introducing a truncation of the underlying normal distribution that generates the observations. In turn, this causes the mean of the data generating process to be less than $\bar{\phi}_k$.

²⁸See Giannitsarou (2006).

For this reason, when using the RGD with the minimum possible t_0 , the initial value for ϕ_k will typically be below the REE, resulting in deviations under learning that are lower than those under rational expectations. Based on outcomes from several numerical experiments, we conjecture that this effect disappears as $\bar{\phi}_k$ moves away from the stationarity bounds -1 and 1.

Finally, if we consider the experiments with CG, we observe that a higher gain coefficient leads to a higher variability of capital in all cases. This can be attributed to a higher variability of the coefficient ϕ_k with a higher gain. This is shown in Figure 4, which displays the evolution of ϕ_k under RLS and CG with ad-hoc initializations that start at $\bar{\phi}_k$. As we see, whereas the variability of ϕ_k is very similar under RLS and CG with $g = 0.02$, it is considerably higher with $g = 0.2$, generating the differences in the variabilities that we have discussed above. Note that this could also explain the lower correlations with output of other variables, such as investment and consumption, that we observe with a higher gain coefficient.

< FIGURE 4 ABOUT HERE >

8.2. The Lucas Tree Model. Table 9 presents some statistics for the Lucas tree model, in particular it reports the standard deviation of the stock price under learning relative to the corresponding standard deviation under RE, using three learning algorithms, RLS, SG and CG-RLS (with gains of 0.02, 0.2 and 0.4). As with the stochastic growth model, whenever DIS is used to initialize the algorithm, the initial value is based on 5 observations.

< TABLE 9 ABOUT HERE >

From this table, we observe the following patterns. First, under RLS or CG-RLS with RGD initializations, we see differences in the relative variability of the stock price, but the numbers are non-monotone with respect to t_0 . Second, as expected, the variability under RLS with DIS initialization is very close to its value under RE, while it is increasing with the gain coefficient under CG-RLS. Third, the relative variability of the price is increasing with the initial value of ϕ with AH initializations, under all different learning algorithms.

< FIGURE 5 ABOUT HERE >

To understand the intuition for these findings, note that the stock price law of motion under learning with our parameterization is given by:

$$p_t = T(\phi_{t-1})d_t = (\beta\phi_{t-1} + 1 - \beta\rho)\rho d_t. \quad (109)$$

This equation shows that the variability of the stock price under learning relative to the variability under rational expectations is increasing in the estimated coefficient ϕ_t . This explains the behavior of the relative variability with respect to the initial value. If the estimated coefficient stays high and above the REE for many periods, then the relative variability is larger than one. Figure 5 shows the evolution of ϕ under RLS and SG with AH

initializations. For example, when RLS starts at $1.2\bar{\phi}$, the coefficient stays above the REE for most periods, this is why the relative variability is higher than one. Moreover, compared to SG starting at the same initial value, the estimated coefficient is overall higher, thus the relative variability is higher, as can be seen in table 9.

8.3. Comparison of the two examples. Our results illustrate some important differences across the two reduced form models considered. In particular, the differences across algorithms seem to be larger in the PFL model, whereas the initialization matters equally for both models.

First, due to the presence of the lag, the FLL reduced form generates a considerable persistence of the coefficient estimate. This persistence will generate less differences between the RLS and SG algorithms. These however are sizeable in the PFL model. Second, the stochastic process has a much higher variance in the second example, implying that all variabilities should be higher in this case. Finally, the effects of potentially higher and more volatile estimates, for example under CG, will considerably be mitigated in the first example due to the projection facility (stationarity condition) imposed on ϕ_k ; this leads to a downward bias that also drives down the variability of the state variable. This is confirmed by the results for the first example with a gain of 0.4, which are not reported here, but show that the variability of the capital is actually lower than under $g = 0.2$. Thus, in this case, imposing the projection facility plays an important role for explaining differences between algorithms. Finally, given our findings from these numerical examples, it is quite clear that the size of the gain in CG learning is critical for determining the dynamics of the system that is studied. Its value should therefore be carefully selected when this algorithm is used in applied work.

9. CONCLUDING REMARKS

In this paper, we provide a self-contained and comprehensive framework for computational implementation of adaptive learning algorithms, in the context of forward looking models with endogenous lags and purely forward looking models. Summarizing our results, the main conclusion of the experiments is that the initial values for learning can be very important for the evolution of the variables of the model. We also find important differences across the algorithms. In particular, a higher gain under constant gain learning can generate a considerably higher variability for the state variable, especially in the purely forward looking reduced form model, where it is not necessary to impose the projection facility. Finally, although the effects of the initial conditions should in theory disappear in the limit, for time horizons that are empirically relevant, the effects often remain and can result in outcomes that are significantly different from those under rational expectations. To reinforce this argument, we simulated the paths of the variables under learning for longer horizons and found that the effects of the initial conditions take a long time to disappear. Therefore, the choice of initial conditions and learning algorithms may be very important if one is interested in using learning in an applied or empirical context.

APPENDIX

A. EXAMPLE 1: Reduced form and Steady State

If we denote the steady state of variable Q_t by \bar{Q} and conventionally set $\bar{Z} = 1$, we obtain:

$$\bar{K} = \theta^{-\frac{1}{1-\alpha}}, \quad (110)$$

$$\bar{C} = \bar{K}^\alpha - \delta\bar{K} = \theta^{-\frac{\alpha}{1-\alpha}} - \delta\theta^{-\frac{1}{1-\alpha}}, \quad (111)$$

$$\bar{Y} = \bar{K}^\alpha = \theta^{-\frac{\alpha}{1-\alpha}}, \quad (112)$$

$$\bar{Z} = 1. \quad (113)$$

To write the reduced form for this example, we start from the expectational equation and eliminate all the variables apart from k and z using the rest of the equilibrium conditions as follows:

$$\begin{aligned} 0 &= E_t [\gamma(c_t - c_{t+1}) - \alpha\beta\theta(1-\alpha)k_t + \alpha\beta\theta z_{t+1}] \\ &= E_t \left[\gamma \left(\frac{\bar{K}\bar{R}}{\bar{C}}k_{t-1} + \frac{\bar{Y}}{\bar{C}}z_t - \frac{\bar{K}}{\bar{C}}k_t - \frac{\bar{K}\bar{R}}{\bar{C}}k_t - \frac{\bar{Y}}{\bar{C}}z_{t+1} + \frac{\bar{K}}{\bar{C}}k_{t+1} \right) - \alpha\beta\theta(1-\alpha)k_t + \alpha\beta\theta z_{t+1} \right] \\ &= E_t \left[\frac{\gamma\bar{K}\bar{R}}{\bar{C}}k_{t-1} + \frac{\gamma\bar{Y}}{\bar{C}}z_t - \frac{\gamma\bar{K}}{\bar{C}}k_t - \frac{\gamma\bar{K}\bar{R}}{\bar{C}}k_t - \frac{\gamma\bar{Y}}{\bar{C}}z_{t+1} + \frac{\gamma\bar{K}}{\bar{C}}k_{t+1} - \alpha\beta\theta(1-\alpha)k_t + \alpha\beta\theta z_{t+1} \right] \\ &= \frac{\gamma\bar{K}\bar{R}}{\bar{C}}k_{t-1} + \frac{\gamma\bar{Y}}{\bar{C}}z_t - \frac{\gamma\bar{K}}{\bar{C}}k_t - \frac{\gamma\bar{K}\bar{R}}{\bar{C}}k_t - \frac{\gamma\bar{Y}}{\bar{C}}E_t z_{t+1} + \frac{\gamma\bar{K}}{\bar{C}}E_t k_{t+1} - \alpha\beta\theta(1-\alpha)k_t + \alpha\beta\theta E_t z_{t+1} \\ &= \frac{\gamma\bar{K}\bar{R}}{\bar{C}}k_{t-1} + \frac{\gamma\bar{Y}}{\bar{C}}z_t - \frac{\gamma\bar{K}}{\bar{C}}k_t - \frac{\gamma\bar{K}\bar{R}}{\bar{C}}k_t - \frac{\gamma\rho\bar{Y}}{\bar{C}}z_t + \frac{\gamma\bar{K}}{\bar{C}}E_t k_{t+1} - \alpha\beta\theta(1-\alpha)k_t + \alpha\beta\theta\rho z_t \\ &= \frac{\gamma\bar{K}}{\bar{C}}E_t k_{t+1} + \frac{\gamma\bar{K}\bar{R}}{\bar{C}}k_{t-1} + \left(\frac{\gamma\bar{Y}}{\bar{C}} - \frac{\gamma\rho\bar{Y}}{\bar{C}} + \alpha\beta\theta\rho \right) z_t - \left(\frac{\gamma\bar{K}}{\bar{C}} + \frac{\gamma\bar{K}\bar{R}}{\bar{C}} + \alpha\beta\theta(1-\alpha) \right) k_t \implies \\ &\left(\frac{\gamma\bar{K} + \gamma\bar{K}\bar{R} + \alpha\beta\theta(1-\alpha)\bar{C}}{\bar{C}} \right) k_t = \frac{\gamma\bar{K}}{\bar{C}}E_t k_{t+1} + \frac{\gamma\bar{K}\bar{R}}{\bar{C}}k_{t-1} + \left(\frac{\gamma\bar{Y} - \gamma\rho\bar{Y} + \alpha\beta\theta\rho\bar{C}}{\bar{C}} \right) z_t. \end{aligned} \quad (114)$$

B. EXAMPLE 2: STEADY STATE

The steady state equations for example two are given by

$$\bar{Z} = 1, \quad \bar{R} = \frac{1-\beta+\delta\beta(1-\tau)}{\beta(1-\tau)}, \quad \bar{W} = (1-\alpha)\left(\frac{\bar{R}}{\alpha}\right)^{\frac{-\alpha}{1-\alpha}}, \quad \bar{Y} = \frac{\bar{R}\bar{K}}{\alpha}, \quad (115)$$

$$\bar{K} = \frac{(\alpha-1)\left(\frac{\bar{R}}{\alpha}\right)^{\frac{-\alpha}{1-\alpha}}\alpha(\xi-1)}{A(\bar{R}-\alpha\delta-\bar{R}\xi+\bar{R}\alpha\xi-\bar{R}\alpha\tau+\alpha\delta\tau)}, \quad \bar{N} = \left(\frac{\bar{R}}{\alpha}\right)^{\frac{1}{1-\alpha}}\bar{K}, \quad (116)$$

$$\bar{C} = \frac{1}{A}(1-\xi)(1-\alpha)\left(\frac{\bar{R}}{\alpha}\right)^{\frac{-\alpha}{1-\alpha}}, \quad G = \bar{K}(\tau(\bar{R}-\delta) + \frac{\xi(1-\alpha)}{\alpha}\bar{R}). \quad (117)$$

REFERENCES

- Bullard, J., Cho, I. K., 2005. Escapist Policy Rules. *Journal of Economic Dynamics and Control* 29, 1841 - 1865.
- Bullard, J., Eusepi, S., 2005. Did the Great Inflation Occur Despite Policy Maker Commitment to a Taylor Rule? *Review of Economic Dynamics* 8, 324 - 359.
- Campbell, J., 1994. Inspecting the Mechanism: An Analytical Approach to the Stochastic Growth Model. *Journal of Monetary Economics* 33, 463-506.
- Carceles-Poveda, E., Giannitsarou, C., 2006. Asset Pricing with Adaptive Learning. Mimeograph.
- Cho, I. K., Williams, N., Sargent, T., 2002. Escaping Nash Inflation. *Review of Economic Studies* 69, 1 - 40.
- Evans, G., Honkapohja, S., 1998. Convergence of Learning Algorithms without a Projection Facility. *Journal of Mathematical Economics* 30, 59 - 86.
- Evans, G., Honkapohja, S., 2001. *Learning and Expectations in Macroeconomics*. Princeton University Press.
- Evans, G., Honkapohja, S., Williams, N., 2006. Generalized Stochastic Gradient Learning. Mimeograph.
- Giannitsarou, C., 2006. Supply-Side Reforms and Learning Dynamics. *Journal of Monetary Economics* 53, 291 - 309.
- Giannitsarou, C., 2005. E-stability Does Not Imply Learnability. *Macroeconomic Dynamics* 9, 276 - 287.
- Ljung, L., Söderström, S., 1986. *Theory and Practice of Recursive Identification*. MIT Press, third printing.
- Marcet, A., Nicolini, J.-P., 2003. Recurrent Hyperinflations and Learning. *American Economic Review* 93, 1476 - 1498.
- Marcet, A., Sargent, T., 1989. Convergence of Least Squares Mechanisms in Self-Referential Linear Stochastic Models. *Journal of Economic Theory* 48, 337 - 368.
- McCallum, B. T., 1983. On Non uniqueness in Linear Rational Expectations Models: An Attempt at Perspective. *Journal of Monetary Economics*, 11, 134 - 168.
- McCallum, B. T., 2006. E-Stability vis-a-vis Determinacy Results for a Broad Class of Linear Rational Expectations Models. *Journal of Economic Dynamics and Control*, forthcoming.
- Milani, F., 2005a. Learning, Monetary Policy Rules, and Macroeconomic Stability. Mimeograph.
- Milani, F., 2005b. Adaptive Learning and Inflation Persistence. Mimeograph.
- Milani, F., 2005c. Expectations, Learning and Macroeconomic Persistence. Mimeograph.
- Orphanides, A., Williams, J., 2005a. The Decline of Activist Stabilization Policy: Natural Rate Misperceptions, Learning and Expectations. *Journal of Economic Dynamics and Control* 29, 1927 - 1950.
- Orphanides, A., Williams, J., 2005b. Inflation Scares and Forecast-Based Monetary

Policy. *Review of Economic Dynamics* 8, 498 - 527.

Sargent, T., 1999. *The Conquest of American Inflation*. Princeton University Press.

Uhlig, H., 1999. A Toolkit for Analyzing Nonlinear Dynamic Stochastic Models Easily, in Marimon, R. and Scott, A. (Eds), *Computational Methods for the Study of Dynamic Economies*, Oxford University Press.

Williams, N., 2003. *Adaptive Learning and Business Cycles*. Mimeograph.

		Relative Deviations			Correlations with y		
		k	i	c	k	i	c
RGD:	$t_0 = \min$	0.1422	1.9580	0.8543	0.5600	0.8630	0.8699
	$t_0 = 15$	0.2423	2.8409	0.4707	0.3982	0.9720	0.8104
	$t_0 = 40$	0.2567	2.9287	0.3774	0.3491	0.9902	0.9031
	$t_0 = 100$	0.2604	2.9823	0.3517	0.3558	0.9918	0.9205
DIS:		0.2774	3.1644	0.4249	0.3406	0.9853	0.6286
REE:		0.2619	3.0101	0.3423	0.3604	0.9917	0.9221

Table 1: **Relative Deviations and Correlations with Output under RLS.** The first four lines show the statistics under RGD with various initial values. The fifth line shows statistics with DIS and the last line shows the statistics under rational expectations.

initial values	$0.900\bar{\phi}_k$	$0.950\bar{\phi}_k$	$\bar{\phi}_k$	$1.010\bar{\phi}_k$	$1.025\bar{\phi}_k$
$0.900\bar{\phi}_z$	0.1815 (0.5254)		0.2586 (0.3326)		0.2927 (0.2919)
$0.950\bar{\phi}_z$		0.2198 (0.4413)			
$\bar{\phi}_z$	0.1939 (0.5285)		0.2780 (0.3418)		0.3161 (0.3058)
$1.010\bar{\phi}_z$				0.2934 (0.3229)	
$1.025\bar{\phi}_z$	0.1970 (0.5292)		0.2828 (0.3441)		0.3215 (0.3091)

Table 2: **Stochastic Growth Model -Relative Deviations and Correlations of Capital with Output under RLS.** The simulations are done with AH initialization. The first row and the first column show the initial values that are used for ϕ_k and ϕ_z . For each initialisation, the first number is the relative deviation and the second number, in parenthesis, is the correlation.

initial values	$0.900\bar{\phi}_k$	$0.950\bar{\phi}_k$	$\bar{\phi}_k$	$1.010\bar{\phi}_k$	$1.025\bar{\phi}_k$
$0.900\bar{\phi}_z$	0.1904 (0.5641)		0.2382 (0.3524)		0.2484 (0.2780)
$0.950\bar{\phi}_z$		0.2243 (0.4772)			
$\bar{\phi}_z$	0.2090 (0.5695)		0.2618 (0.3604)		0.2732 (0.2866)
$1.010\bar{\phi}_z$				0.2691 (0.3328)	
$1.025\bar{\phi}_z$	0.2137 (0.5708)		0.2677 (0.3623)		0.2793 (0.2887)

Table 3: **Stochastic Growth Model - Relative Deviations and Correlations of Capital with Output under SG.** The simulations are done with AH initialization. The first row and the first column show the initial values that are used for ϕ_k and ϕ_z . For each initialisation, the first number is the relative deviation and the second number, in parenthesis, is the correlation.

initial values	$0.900\bar{\phi}_k$	$0.950\bar{\phi}_k$	$\bar{\phi}_k$	$1.010\bar{\phi}_k$	$1.025\bar{\phi}_k$
$0.900\bar{\phi}_z$	0.1910 (0.5535)		0.2485 (0.3394)		0.2711 (0.2667)
$0.950\bar{\phi}_z$		0.2250 (0.4637)			
$\bar{\phi}_z$	0.2070 (0.5579)		0.2706 (0.3477)		0.2958 (0.2767)
$1.010\bar{\phi}_z$				0.2822 (0.3211)	
$1.025\bar{\phi}_z$	0.2110 (0.5590)		0.2762 (0.3497)		0.3020 (0.2792)

Table 4: **Stochastic Growth Model - Relative Deviations and Correlations of Capital with Output under CG-RLS.** The simulations are done with AH initialization and gain $g = 0.02$. The first row and the first column show the initial values that are used for $\bar{\phi}_k$ and $\bar{\phi}_z$. For each initialisation, the first number is the relative deviation and the second number, in parenthesis, is the correlation.

initial values	$0.900\bar{\phi}_k$	$0.950\bar{\phi}_k$	$\bar{\phi}_k$	$1.010\bar{\phi}_k$	$1.025\bar{\phi}_k$
$0.900\bar{\phi}_z$	0.2272 (0.3412)		0.3151 (0.3289)		0.3436 (0.3284)
$0.950\bar{\phi}_z$		0.2671 (0.3386)			
$\bar{\phi}_z$	0.2339 (0.3494)		0.3270 (0.3389)		0.3567 (0.3389)
$1.010\bar{\phi}_z$				0.3397 (0.3396)	
$1.025\bar{\phi}_z$	0.2352 (0.3512)		0.3992 (0.3417)		0.3610 (0.3423)

Table 5: **Stochastic Growth Model - Relative Deviations and Correlations of Capital with Output under CG-RLS.** The simulations are done with AH initialization and gain $g = 0.2$. The first row and the first column show the initial values that are used for $\bar{\phi}_k$ and $\bar{\phi}_z$. For each initialisation, the first number is the relative deviation and the second number, in parenthesis, is the correlation.

		Relative Deviations			Correlations with y		
		k	i	c	k	i	c
RGD:	$t_0 = \min$	0.1507	2.1240	0.8412	0.6102	0.8330	0.8395
	$t_0 = 15$	0.2454	2.8510	0.4510	0.3839	0.9748	0.8294
	$t_0 = 40$	0.2616	2.9694	0.3694	0.3414	0.9894	0.8845
	$t_0 = 100$	0.2673	3.0384	0.3394	0.3454	0.9912	0.8943
DIS:		0.2689	3.0593	0.3582	0.3449	0.9908	0.7890
REE:		0.2619	3.0101	0.3423	0.3604	0.9917	0.9221

Table 6: **Relative Deviations and Correlations with Output under CG-RLS**, $g = 0.02$. The first four lines show the statistics under RGD with various initial values. The fifth line shows statistics with DIS and the last line shows the statistics under rational expectations.

		Relative Deviations			Correlations with y		
		k	i	c	k	i	c
RGD:	$t_0 = \min$	0.1936	2.3162	0.8307	0.3174	0.8419	0.7871
	$t_0 = 15$	0.2613	3.0630	0.7363	0.3553	0.8734	0.5774
	$t_0 = 40$	0.3021	3.4650	0.7093	0.3482	0.8854	0.4324
	$t_0 = 100$	0.3157	3.5853	0.7012	0.3417	0.8916	0.3820
DIS:		0.3265	3.6859	0.7130	0.3398	0.8944	0.3427
REE:		0.2619	3.0101	0.3423	0.3604	0.9917	0.9221

Table 7: **Relative Deviations and Correlations with Output under CG-RLS**, $g = 0.2$. The first four lines show the statistics under RGD with various initial values. The fifth line shows statistics with DIS and the last line shows the statistics under rational expectations.

	Initial values	Rel. deviations	Percentage of periods below REE	
			ϕ_k	ϕ_z
RLS-RGD	$t_0 = \min$	0.1422	82	68
	$t_0 = 15$	0.2423	66	58
	$t_0 = 40$	0.2567	57	61
	$t_0 = 100$	0.2604	57	56
RLS-DIS		0.2774	50	51
RLS-AH	$1.025\bar{\phi}$	0.3215	21	34
	$1.010\bar{\phi}$	0.2934	35	45
	$\bar{\phi}$	0.2780	50	51
	$0.950\bar{\phi}$	0.2198	99	71
	$0.900\bar{\phi}$	0.1815	99	83
SG	$1.025\bar{\phi}$	0.2793	0	0
	$1.010\bar{\phi}$	0.2691	0	0
	$\bar{\phi}$	0.2618	65	57
	$0.950\bar{\phi}$	0.2243	100	100
	$0.900\bar{\phi}$	0.1904	100	100
REE		0.2618		

Table 8: **Percentage of Periods Below the REE for the Stochastic Growth Model.** The simulations are based on 2000 replications. The first two columns show the initialization method and the specifications for the initial value. The third column shows the average relative deviations and the two last columns show the average percentage number of periods for which ϕ_k and ϕ_z is below the REE.

	RGD				DIS	AH		
	min	15	40	100		$0.8\bar{\phi}$	$\bar{\phi}$	$1.2\bar{\phi}$
RLS	0.8222	1.0782	0.9449	1.0029	1.0301	0.8706	1.0308	1.1911
SG						0.8132	1.0006	1.1880
CG-RLS, $g = 0.02$	0.8687	1.0757	0.9521	1.0113	1.0052	0.8334	1.0098	1.1861
CG-RLS, $g = 0.20$	1.1166	1.2233	1.1374	1.1721	1.1621	1.0507	1.1614	1.2741
CG-RLS, $g = 0.40$	1.9243	1.4612	1.3888	1.4149	1.3995	1.3148	1.3995	1.4870

Table 9: **Lucas Tree Model.** Deviations of the stock price under learning, relative to deviations under rational expectations. The statistics are based on simulations under RLS, SG and CG-RLS learning. The first and second rows show the various initialization methods.

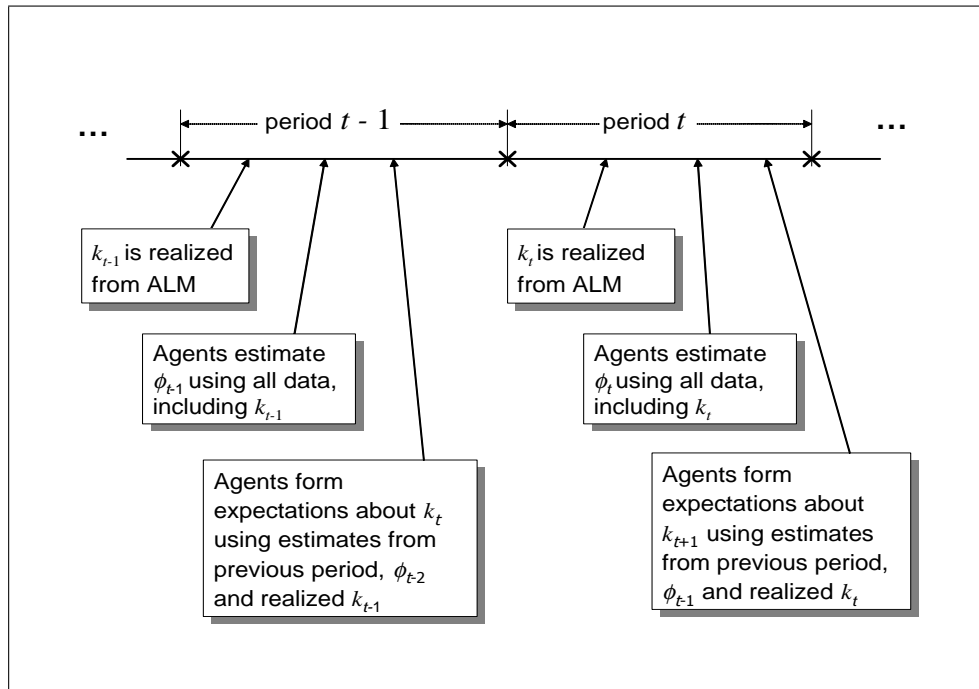


Figure 1: **Timeline of adaptive learning.**

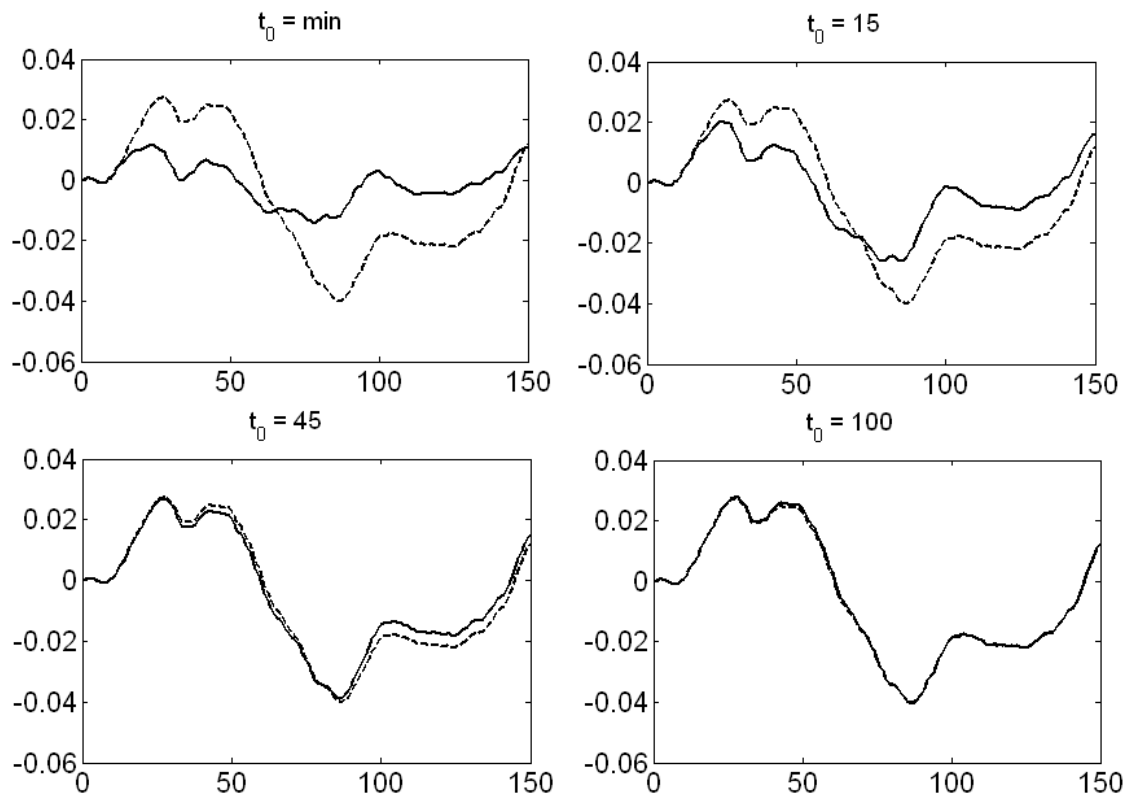


Figure 2: **Evolution of k_t** . The solid lines show simulations of capital in the growth model, under RLS with RGD. The dashed lines show evolution of capital under rational expectations.

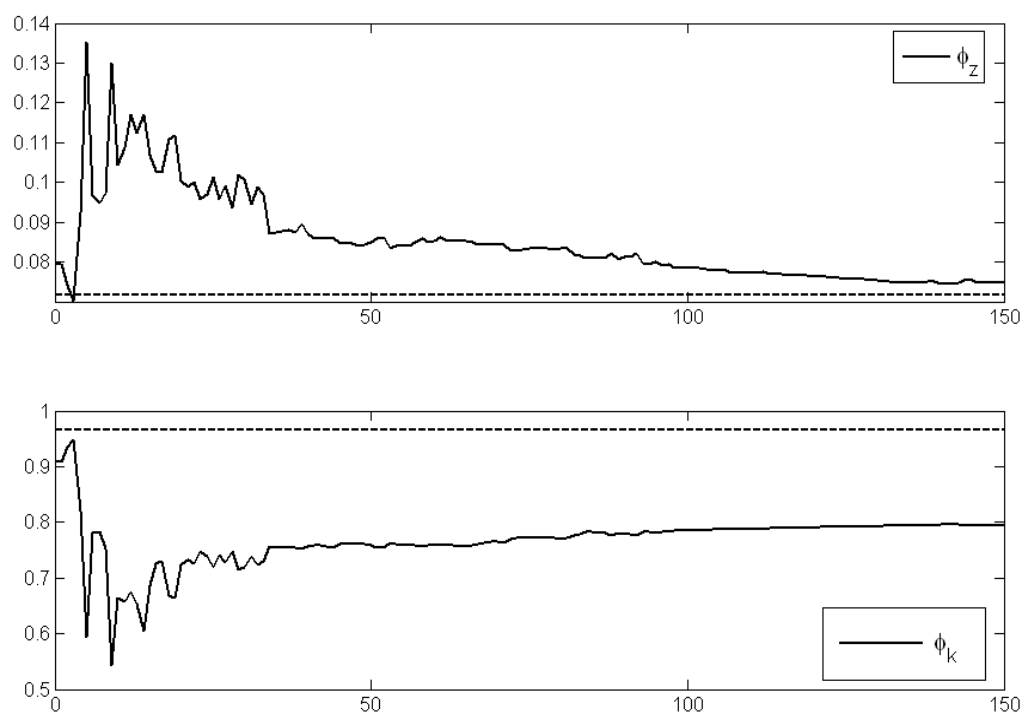


Figure 3: **Evolution of ϕ .** Simulation of the coefficients for the growth model under RLS, using RGD with the minimum possible t_0 .

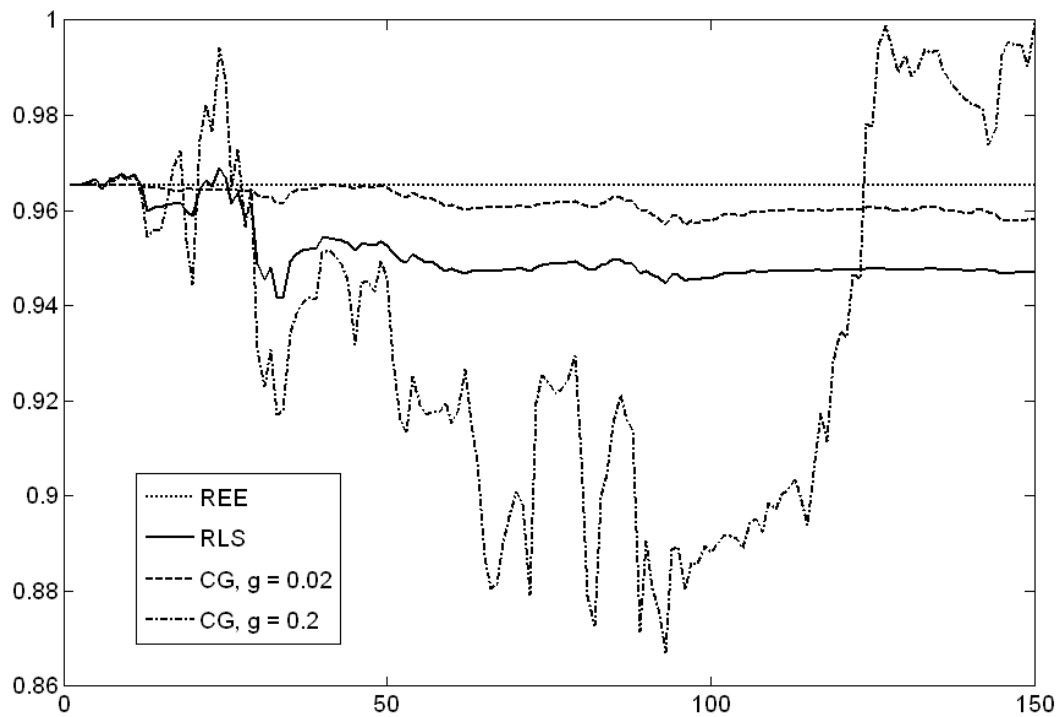


Figure 4: **Evolution of ϕ_k .** Comparison of dynamics under RLS (solid line) and CG (dashed line, $g = 0.02$ and dash-dotted line, $g = 0.2$). The initial value is $\bar{\phi}_k$. The REE is shown with the dotted straight line.

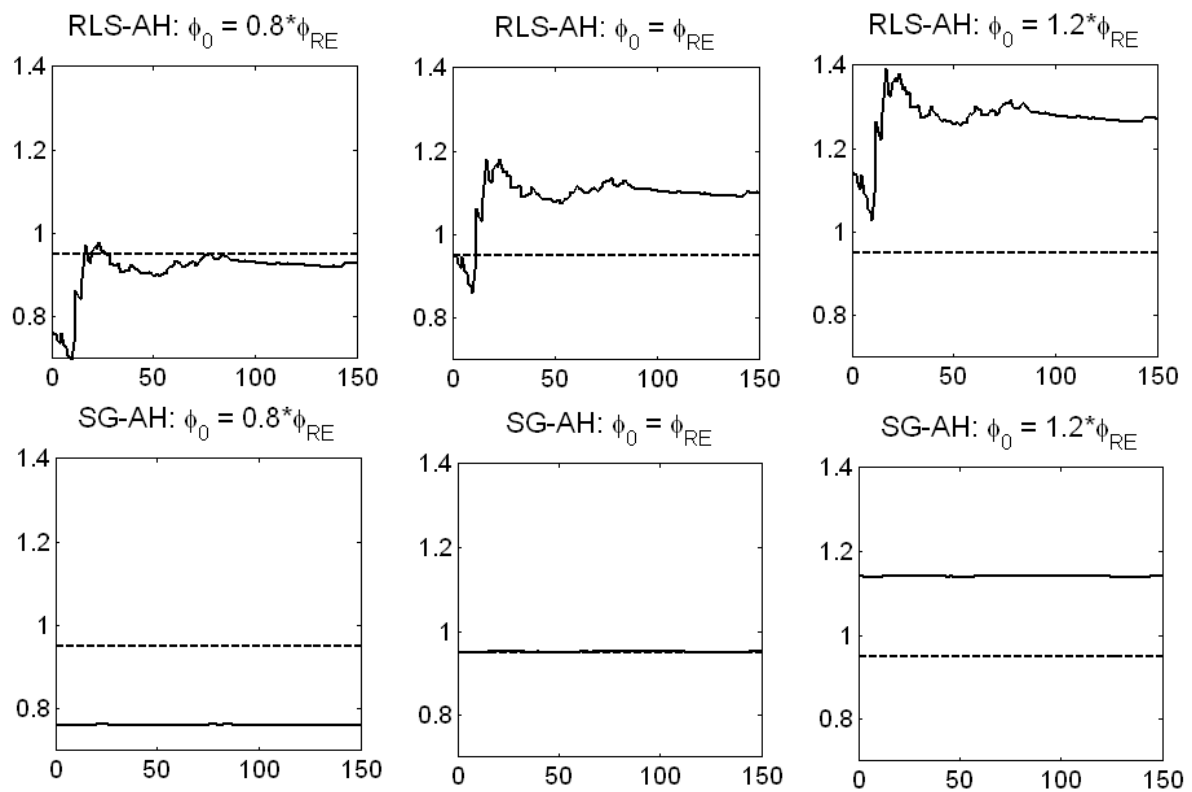


Figure 5: **Comparison of RLS and SG.** The first row of graphs shows evolution of ϕ in the Lucas model, under RLS with three initial values. The second row shows the evolution of ϕ under SG. The dashed line is the REE.