

2012

Application of an Adaptive Step-Size Algorithm in Models of Hyperinflation

Olena Kostyshyna
Portland State University

Let us know how access to this document benefits you.

Follow this and additional works at: https://pdxscholar.library.pdx.edu/econ_fac



Part of the [Macroeconomics Commons](#)

Citation Details

Olena Kostyshyna (2012). Application of an Adaptive Step-Size Algorithm in Models of Hyperinflation. *Macroeconomic Dynamics*, 16, pp 355-375.

This Article is brought to you for free and open access. It has been accepted for inclusion in Economics Faculty Publications and Presentations by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

APPLICATION OF AN ADAPTIVE STEP-SIZE ALGORITHM IN MODELS OF HYPERINFLATION

OLENA KOSTYSHYNA

Portland State University

An adaptive step-size algorithm [Kushner and Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed., New York: Springer-Verlag (2003)] is used to model time-varying learning, and its performance is illustrated in the environment of Marcet and Nicolini [*American Economic Review* 93 (2003), 1476–1498]. The resulting model gives qualitatively similar results to those of Marcet and Nicolini, and performs quantitatively somewhat better, based on the criterion of mean squared error. The model generates increasing gain during hyperinflations and decreasing gain after hyperinflations end, which matches findings in the data. An agent using this model behaves cautiously when faced with sudden changes in policy, and is able to recognize a regime change after acquiring sufficient information.

Keywords: Time-Varying Gain, Adaptive Expectations, Hyperinflation, Learning in Macroeconomics

1. INTRODUCTION

This paper introduces a new algorithm to model learning in an environment of hyperinflations. Explanations of hyperinflation using rational expectations rely on bubble equilibria [Sargent and Wallace (1987)], and recurrence of hyperinflations is explained by sunspots [Funke et al. (1994)]. Other explanations employ departures from rational expectations and introduce learning [Marcet and Nicolini (2003), Adam et al. (2006), Sargent et al. (2006)]. These latter use learning mechanisms that differ in the specification of the speed of adjustment to new information.

The speed of adjustment to new information is called the speed of update or gain. The gain determines how quickly the learning mechanism reacts to new observations, and how much weight is given to new relative to old information. Two common assumptions about the gain used in the literature are least-squares learning and constant-gain (perpetual) learning. In least-squares learning, the gain decreases with time and gives equal weight to all past observations. If an agent knows that he

I am very grateful to Ken Kasa and Jasmina Arifovic for helpful discussions and comments on this paper. And I would like to thank the anonymous referee for helpful comments and suggestions. Address correspondence to: Olena Kostyshyna, Department of Economics, Portland State University, 1721 SW Broadway, Portland, OR 97201, USA; e-mail: okostysh@pdx.edu.

is in a stationary environment, he will use a decreasing gain. Examples of research using least-squares learning include Evans and Honkapohja (2001), and Marcat and Sargent (1989a, 1989b). In general, least-squares learning can converge to a rational expectations equilibrium under certain conditions [Bullard (1991, 2006), Evans and Honkapohja (2001), Bullard and Mitra (2002)].

In contrast, a constant-gain algorithm discounts past data and gives more weight to recent observations. A constant-gain algorithm is thus better at tracking structural changes. If an agent believes that the environment is not stationary, and that regime changes are possible, constant-gain learning is used. Constant-gain learning can be interpreted as agents not trusting the government, and so being alert to possible changes in the government's policy. Once change happens, an agent using a constant gain is ready to view it as a regime switch and updates his estimates accordingly. Evans and Honkapohja (1993, 2001), Sargent (1999), Orphanides and Williams (2005), Chakraborty and Evans (2008), and Cho and Kasa (2008) are several examples of research using constant-gain learning. Sargent (1999) finds that least-squares learning converges to a Nash equilibrium in the Kydland–Prescott model [Kydland and Prescott (1977)]. Sargent (1999) suggests using constant-gain learning—it can allow a government using a misspecified model to achieve a superior Ramsey outcome and escape from Nash inflation. In a changing environment, constant-gain learning does not converge to a rational expectations equilibrium, but it may converge to an ergodic distribution around a REE [Sargent (1999), Evans and Honkapohja (2001)].

Several questions about constant-gain learning remain to be addressed. First is the choice of the gain parameter. In most of the literature, this parameter is chosen in an ad hoc manner to produce desired properties in the model at hand. The value of the gain is generally quite important for the results. The recent literature estimates the constant gain either from the data or from surveys of professional forecasters [Orphanides and Williams (2004), Branch and Evans (2005), Sargent et al. (2006), Milani (2007, 2008)].

A second question is whether constant-gain beliefs can be self-confirming. The stated motivation for using a constant-gain algorithm is a nonstationary environment, but constant-gain learning is often used in models that are stationary. In these models, if agents were to use decreasing gain, their expectations would be validated just as well as they are validated when they use constant-gain expectations [Chakraborty and Evans (2008)].

A third issue is that one would expect agents to learn at different speeds depending on the economic environment. Cagan (1956) uses adaptive expectations to study hyperinflation and points out the limitations of this scheme: adaptive mechanisms with constant gain constrain agents to adjust their forecasts by a constant proportion of their forecast error.¹ Cagan (1956) estimates gains during hyperinflation and finds that agents update more slowly in the earlier periods of hyperinflation, and more quickly in the later periods of hyperinflation. Khan (1977) finds support for variable gain in the data as well—the update speed increases with the variability of inflation. Khan (1977) specifies the gain as a function of

the absolute level of inflation (using Cagan's finding that the gain increases as hyperinflation unfolds) and the variability of inflation, measured by the absolute change of inflation. A similar specification for a variable gain is derived in a model with rational expectations by Mussa (1981). Silveira (1973) finds support for increasing gain in hyperinflation data from Brazil. These papers point out that agents need to revise their expectations more quickly, and so use a higher gain, because slow adjustment to forecast errors can be costly in hyperinflationary environments.

There are more recent examples using time-varying gain parameters. Marcet and Nicolini (2003) introduce a learning mechanism that combines a constant gain and a decreasing gain to model recurrent hyperinflation in Latin America. Their agent switches from a decreasing gain to a constant gain when forecast errors rise above a critical level, and then returns to a decreasing gain when forecast errors fall below the critical level. The value of the constant gain is chosen so that the learning algorithm satisfies the lower bound on rationality. Timmermann (1993) models learning in an environment with infrequent structural changes, where the timing of the regime shifts is known to the agent. The agent uses a decreasing gain when the structure is unchanged and a constant gain when a structural break occurs. Milani (2007) allows for a structural break that changes the value of constant gain. Evans and Ramey (2005) derive the Nash equilibrium gain and show that agents using a recursive prediction error (RPE) algorithm [Ljung and Soderström (1983)] are able to adjust their gain to an equilibrium value that varies with policy changes.

In this paper, I introduce a new time-varying gain algorithm. I use an adaptive-step-size algorithm as described in Benveniste et al. (1990), Kushner and Yang (1995), and Kushner and Yin (2003, pp. 69–73). The successful use of a constant-gain algorithm depends on the choice of the gain. The optimal choice of the gain depends on the relative variability of the time-varying parameters to be estimated, the probability distributions of the data, and the observation noise. If the probability distributions of the data and parameters vary greatly, then a higher gain is optimal. However, if observation noise is high, then a smaller gain is better. The adaptive step-size algorithm allows the gain to evolve in response to changes in the environment. Estimation consists of two parts: (i) estimation of the model parameters, and (ii) estimation of the gain (the details are provided in the section "Specification of expectations"). This algorithm is commonly used in engineering applications, and should be useful in economic learning models. Its economic interpretation is that agents adjust the speed of their learning based on recent experience. This procedure thus avoids the problems of committing to a single value for the gain.

To illustrate the performance of the adaptive step-size algorithm, I apply it to the environment in Marcet and Nicolini (2003), in which an agent learns about forming inflation expectations. The agent does not know whether there are regime changes or not, and adapts his forecasts and learning speed based on his observations using the adaptive step-size algorithm.

1.1. Main Findings

The adaptive step-size algorithm specifies a learning scheme that is endogenous to the model and to changes in policy. This addresses the criticism that boundedly rational mechanisms are exogenous to the model. This algorithm provides a universal approach to how gain parameters change in response to changes in the model environment, and avoids the arbitrariness of gain parameter specifications in the previous literature.

Interestingly, the simulations show that the model in this paper produces behavior of inflation and inflation expectations qualitatively similar to that in Marcet and Nicolini's (2003) model (MN below). Mean squared errors are similar to those of Marcet and Nicolini (2003). This suggests that adaptive step-size algorithms can provide a natural and convenient method for incorporating a state-dependent learning speed. Previous work along these lines has had to resort to ad hoc mechanisms.

An adaptive step-size algorithm also produces interesting dynamics in the gain. The value of gain increases as hyperinflation develops. The increasing gain during hyperinflations matches the empirical findings in Cagan (1956), Silveira (1973), and Khan (1977). The adaptive step-size learning specification matches findings of increasing speed of update during hyperinflation better than the mechanism in MN. In the MN model, the agent switches from a decreasing gain to a constant gain at the beginning of hyperinflation and continues to use a constant gain during hyperinflation. In my model, the agent revises his or her update speed optimally based on the adaptive step-size algorithm.

In my model, the gain behaves differently than in MN after a hyperinflation ends with the implementation of an exchange rate rule (ERR). In MN, the agent continues to update using constant gain. In my model, the agent switches to a *lower* gain right after an ERR is implemented. The decrease in gain has the following behavioral interpretation. The agent does not know that a regime change has happened and inflation will be low, and/or does not believe that the new policy regime is credible. The agent has repeatedly underpredicted inflation during hyperinflation. When inflation drops, the agent does not discard past experience quickly, and does not rush to revise down his inflation forecasts, because he does not believe that a new regime has started or will prevail successfully.

This behavior of the gain (increases during hyperinflation, decreases after implementation of the ERR) is similar to the mechanism modeled in Cho and Sargent (1997). They specify a learning algorithm in which agents are skeptical that the government can stick to good economic policy (p. 10). When applied to hyperinflationary environments, this mechanism implies the following. During a hyperinflation, agents update their beliefs by putting more weight on recent data, because they understand that the government does not use good economic policy. When agents observe a sudden drop in inflation, they update by placing less weight on recent data, because they do not expect the government to maintain good policy.

I estimate the model for Argentina, Bolivia, Brazil, and Peru using the simulated method of moments with a genetic algorithm. Simulations for the estimated parameters exhibit hyperinflation for all countries except Brazil. This suggests that a more sophisticated estimate (for example, based on transition probabilities for inflation) may be needed.

1.2. Related Literature

Sargent and Wallace (1987) show that a high-inflation steady state (with perverse comparative statics) is stable under perfect foresight and explain hyperinflations as rational bubble equilibria. Marcet and Sargent (1989b) study the model in Sargent and Wallace (1987) with least-squares learning, and find that the low-inflation steady state is stable under least-squares learning, whereas the high-inflation steady state is not. Adam et al. (2006) show that hyperinflationary paths near the high-inflation steady state are stable if agents use contemporaneous data.

Marcet and Nicolini (2003) study the Sargent and Wallace (1987) model (with stochastic i.i.d. seignorage) by introducing an endogenous learning scheme that combines least-squares learning and constant-gain learning. One of the mechanisms is used depending on the size of forecast error. Marcet and Nicolini (2003) are able to address stylized facts of recurrent hyperinflations in Latin American countries.

Sargent et al. (2006) estimate the same model with constant-gain learning and some modifications (e.g., specification of seignorage as a Markov switching process) for Latin American countries that experienced hyperinflation. They conclude that changes in inflation can be attributed to the learning dynamics (switches of perceived inflation between low and high self-confirming equilibria) and/or to changes in fundamentals (change in seignorage).

1.3. Organization

I describe the environment and the expectation-formation mechanism in Section 2. The results of the simulations and analysis are presented in Section 3. Section 4 presents the results of estimation of model and is followed by the conclusion.

2. MODEL

I study a model that consists of a money demand equation, a government budget constraint, an exogenous process for seignorage, and a specification of expectation formation that is different from rational expectations. Marcet and Nicolini (2003) and Sargent et al. (2006) use the same model with different expectation mechanisms.

The demand for money is given by a Cagan-style specification:

$$\frac{M_t}{P_t} = \phi - \gamma \phi \frac{P_{t+1}^e}{P_t}, \quad (1)$$

where $\phi, \gamma > 0$ are parameters, M_t is nominal balances as a percent of output at time t , P_t is the price level at time t , and P_{t+1}^e is the expected price level for time $t + 1$.

The government supplies money to finance seignorage. If inflation is above a certain critical level, the government implements an ERR. If there is no need for the ERR, then the government budget constraint is

$$M_t = M_{t-1} + d_t P_t, \tag{2}$$

where the seignorage process d_t is i.i.d. I follow the assumption in Marcet and Nicolini (2003) that seignorage is normally distributed, $N(E(d), \sigma_d^2)$, $\sigma_d = 0.01$, and truncated to have positive values. The equilibrium values of nominal balances and prices $\{M_t, P_t\}_{t=0}^\infty$ are determined from (1), (2), and an expectation formation equation. I denote the inflation rate as $\pi_t = P_t/P_{t-1}$. Under rational expectations, there are two deterministic steady state inflation rates for $d = E(d_t)$:

$$\pi_{1,2} = \frac{1 + \gamma - d/\phi \pm \sqrt{(1 + \gamma - d/\phi)^2 - 4\gamma}}{2\gamma}. \tag{3}$$

When inflation exceeds the upper critical level $\pi > \pi^U$, or there is no positive price level that clears the market without adjustment of foreign reserves, the government imposes an ERR. From purchasing power parity (PPP), the exchange rate is

$$e_t = \frac{P_t}{P_t^f}, \tag{4}$$

where P_t is the domestic price level, and P_t^f is the foreign price level. Using PPP, we can write the domestic inflation rate as

$$\frac{P_t}{P_{t-1}} = \frac{e_t P_t^f}{e_{t-1} P_{t-1}^f}. \tag{5}$$

The ERR sets the exchange rate in (5) so that the inflation rate is equal to the targeted rate $\bar{\pi}$:

$$\frac{P_t}{P_{t-1}} = \bar{\pi}. \tag{6}$$

The targeted rate $\bar{\pi}$ is chosen to be equal to the low-inflation steady state in (3). Under ERR, the equilibrium price level is determined from (6). Money demand is determined from (1) together with the inflation expectation specification. If money demand in (1) is not equal to money supply in (2), the government adjusts foreign reserves so that the money supply satisfies

$$M_t = M_{t-1} + d_t P_t + e_t (R_t - R_{t-1}). \tag{7}$$

2.1. Expectation Formation

Agents form their expectations of inflation, $\beta_{t+1} = P_{t+1}^e / P_t$, adaptively as:

$$\beta_t = \beta_{t-1} + a_t(\pi_{t-1} - \beta_{t-1}) \tag{8}$$

for some initial value β_0 . Thus inflation forecasts are revised by the last forecast error weighted by the gain a_t .

In this paper, the gain a_t evolves based on the adaptive step-size algorithm suggested in Benveniste et al. (1990) and Kushner and Yang (1995) and described in Kushner and Yin (2003, p. 71). This algorithm specifies the endogenous behavior of the gain in response to developments in the economy. The algorithm can therefore be thought of as providing a unified specification of how gains evolve over time, and so avoiding an ad hoc mechanism of discrete changes in the gain in response to big forecast errors, as in Marcet and Nicolini (2003). In my model, the gain follows

$$a_t = \prod_{[a_-, a_+]} [a_{t-1} + \mu(\pi_{t-1} - \beta_{t-1})V_{t-1}], \tag{9}$$

$$V_t = V_{t-1} - a_{t-1}V_{t-1} + (\pi_{t-1} - \beta_{t-1}), V_0 = 0. \tag{10}$$

In (9), μ is the step size in the stochastic approximation of the a_t process (loosely speaking, μ is “gain on the gain”). $\prod_{[a_-, a_+]}$ is a projection operator that sets the gain a_t equal to a_- when it falls below this value, and sets the gain equal to a_+ when it rises above this value. Kushner and Yin (2003) show that the performance of a nonadaptive algorithm (with constant gain) is much more sensitive to the choice of constant gain a than the adaptive step-size algorithm is to the choice of the step size μ . Although it is necessary that $0 < \mu \ll a_-$ for the proofs, the lower bound a_- is not so important in applications. However, the upper bound a_+ is very important for the performance, and is often chosen close to the point where the algorithm becomes unstable.

In (10), V_t denotes the “derivative” of the estimated parameter (β in this case) with respect to the gain a for the stationary process. The process β_t is not a classical function of a , but its distribution depends on a . Kushner and Yang (1995) interpret V_t as the desired derivative [also see Kushner and Yin (2003)].

Kushner and Yin (2003) indicate that the gain should be chosen to minimize MSE, so one should adjust the gain in such a way as to ensure that $d(E(\text{FE})^2) < 0$, where $\text{FE} = \pi_t - \beta_t$ is the forecast error. Because $d(E(\text{FE})^2) = \frac{\partial}{\partial a}(E(\text{FE})^2)da$, and

$$\frac{\partial}{\partial a}(E(\text{FE})^2) = -2(\text{FE})\frac{\partial \beta}{\partial a}, \tag{11}$$

one should choose da so that

$$-2(\text{FE})\frac{\partial \beta}{\partial a}da < 0, \tag{12}$$

that is, da should have the same sign as $(FE) \frac{\partial \beta}{\partial a}$. So, letting μ be small and positive, we set

$$da = \mu(FE) \frac{\partial \beta}{\partial a}. \quad (13)$$

Finally, because

$$\beta_t = \beta_{t-1} + a(\pi_{t-1} - \beta_{t-1}), \quad (14)$$

we may compute $\frac{\partial \beta}{\partial a}$ to get

$$\frac{\partial \beta_t}{\partial a} = (1 - a) \frac{\partial \beta_{t-1}}{\partial a} + FE_{t-1}. \quad (15)$$

Combining (13) and (15) yields equations (9) and (10).²

The intuition behind this learning mechanism is as follows. The change in gain (9) is driven by the discounted past errors, V , and the last-period forecast error, $\pi_t - \beta_t$.³ When the last-period forecast error is in the same direction as the discounted past errors, the agent increases the gain. This means that if the agent keeps making the same error, he wants to increase the response to the last-period forecast error when updating forecasts. If the last-period forecast error is in a different direction from the discounted past errors, the agent decreases the gain. This means that when the agent encounters something contradictory to his past experience, the agent reduces the response to forecast errors until he learns more. The value of V (10) depends on the size of the last-period forecast error $(\pi_{t-1} - \beta_{t-1})$ relative to the size of past discounted errors $(V_{t-1} - a_{t-1}V_{t-1})$. For example, if the latest forecast error is small relative to the past discounted errors, then V changes slowly.

3. SIMULATION RESULTS

I simulate the economy using the same parameter values as in Marcet and Nicolini (2003).⁴ The parameters in the money demand equation are $\gamma = 0.4$ and $\phi = 0.37$, and mean seignorage is $E(d) = 0.049$. The model has two steady states that become closer for higher values of average seignorage. For the above money demand parameters, the maximum value of average seignorage for which a rational expectations equilibrium exists is $E(d) = 0.05$. When mean seignorage is closer to its maximum value, it is easier for the system to move above the high-inflation steady state and to explode into a hyperinflation.

The parameters of the step-size algorithm are $\mu = 0.001$, $a_- = 0.01$, and $a_+ = 0.6$. The initial value of the gain is 0.2.⁵ It is important to note that this upper boundary for the gain is not binding, as in the simulations, the gain rarely reached values equal to or above it. I choose the initial expectation as $\beta_0 = \pi_0 = \pi_{\text{low}}$, equal to the initial inflation started at the low inflation steady state. In the replication of Marcet and Nicolini (2003), agents start with the least-squares rule,⁶ and the gain in period $t = 1$ is equal to $\frac{1}{t} = 1$.

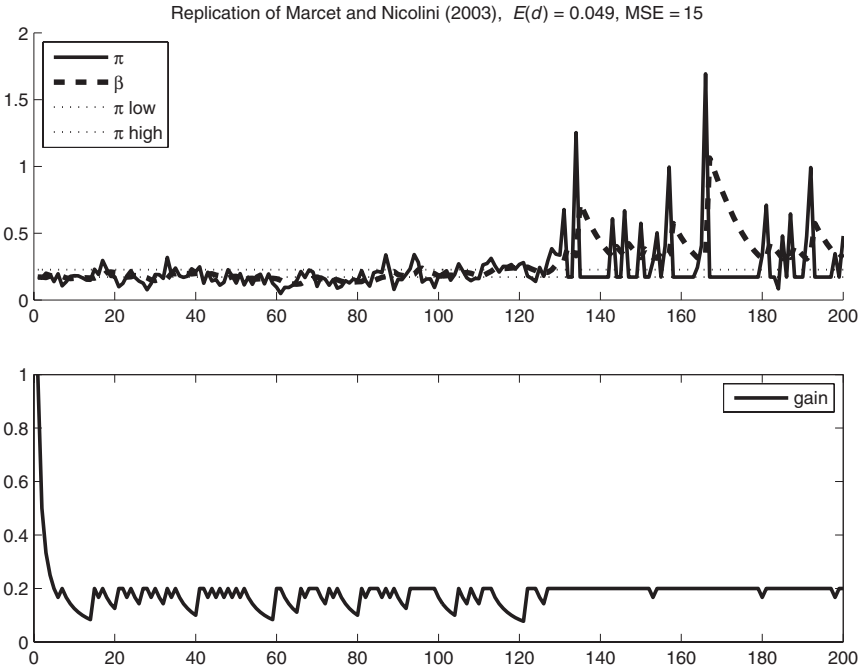


FIGURE 1. Replication of the model by Marcet and Nicolini (2003). The first panel shows actual inflation, π , and expected inflation, β . The second panel shows gain.

I will compare the simulations based on Marcet and Nicolini (2003) and the simulations based on the gain specification in (9). Figure 1 presents a replication of Marcet and Nicolini's (2003) model. Figures 2 and 3 present the results of simulations with the adaptive step-size algorithm.

Figures 1 and 2 are based on the same seed; the only difference is the specification of the gain. These two figures look very similar in terms of the behavior of actual inflation (solid line) and inflation expectations (dashed line). For example, the timing and magnitude of inflation are similar. Expectations of inflation adapt slowly after the end of hyperinflation in both models. The key difference between the two models is in how agents revise their inflation expectations; i.e., the behavior of the gain is different.

Hyperinflation develops as a result of high seignorage shocks that lead to higher inflation. Higher actual inflation leads to higher expected inflation. When expected inflation is above the high steady state, actual inflation increases even more in response to increases in expected inflation. This further raises expected inflation, so that hyperinflation can erupt. As hyperinflation develops, the agent using an adaptive step-size algorithm increases the gain. The agent using an adaptive step-size algorithm can adjust the gain flexibly as the economic environment changes. As hyperinflation unfolds, the agent's inflation forecasts are repeatedly below

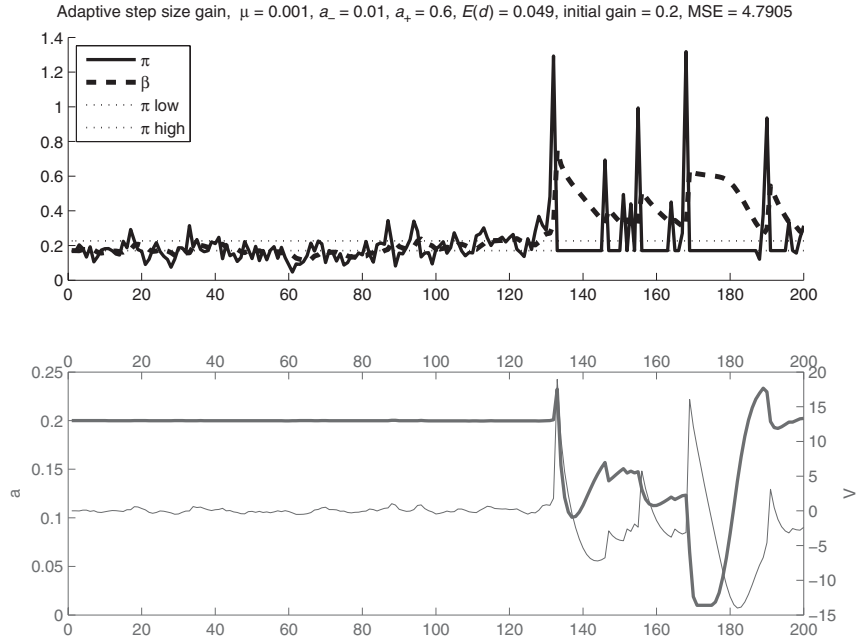


FIGURE 2. Typical simulation with the adaptive step-size algorithm. The first panel shows actual inflation, π , and expectations of inflation, β . The second panel shows adaptive step-size gain on the left y-axis (thick line) and the derivative V on the right y-axis (thin line).

actual inflation. Therefore, the agent begins to increase the gain so that his inflation expectations catch up to actual inflation. A higher gain means that the agent updates his inflation expectations with higher responses to forecast errors. If the forecast errors are positive (as is the case during hyperinflation), increasing the gain means that the agent increases his inflation expectation by a higher proportion of forecast error. The increase in gain speeds up the update of expected inflation and, thus, contributes to the development of hyperinflation.

Cagan (1956) estimates a model with adaptive expectations for subperiods of hyperinflation and finds that the gain increases as hyperinflation evolves. The value of gain increases in the regression when the initial periods of hyperinflation are dropped. Silveira (1973) applies the same estimation procedure for Brazil [not considered in Cagan (1956)]. Silveira (1973) finds that the gain increases in regressions where the later periods of hyperinflation are added, and the gain keeps increasing if the initial periods of hyperinflation are dropped. Khan (1977) specifies the gain as a function of inflation and inflation variability and finds that the gain is positively related to inflation variance. Khan's specification is related to Mussa (1981), who finds that update speed is a function of inflation variance in a model with rational agents. In this paper, the agent revises his speed of

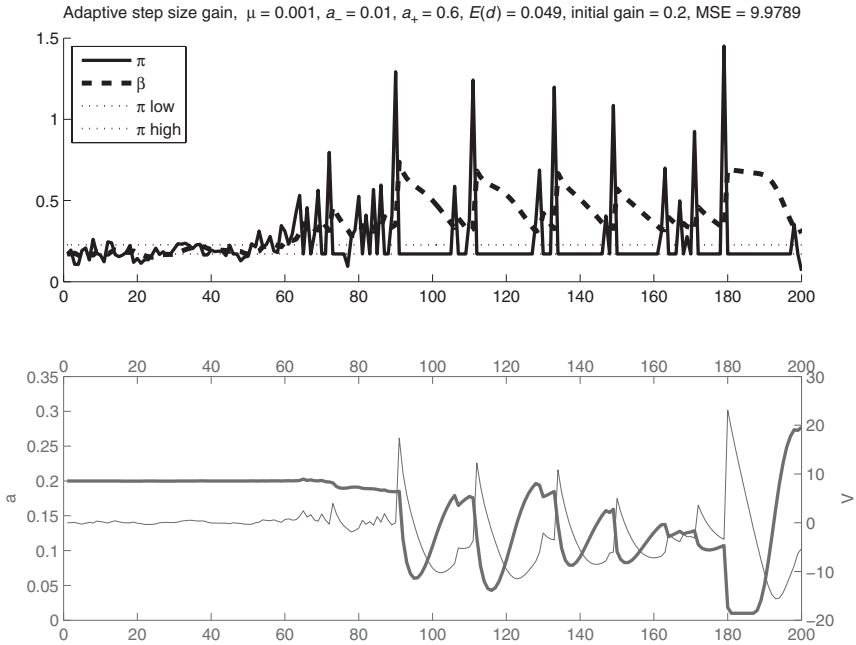


FIGURE 3. Typical simulation with the adaptive step-size algorithm. The first panel shows actual inflation, π , and expectation of inflation, β . The second panel shows adaptive step-size gain on the left y-axis (thick line) and the derivative V on the right y-axis (thin line).

update optimally based on the adaptive step-size algorithm (9), and the resulting behavior of the gain matches these empirical findings of increasing gain during hyperinflation: the gain continues to increase as hyperinflation develops.

The adaptive step-size algorithm captures an increasing speed of update during hyperinflation better than the mechanism in MN. In this paper, the gain keeps increasing as hyperinflation develops. In the MN model, as hyperinflation progresses, the agent makes forecast errors that are higher than the critical level, and therefore, the agent switches from a decreasing gain to a constant gain of 0.2. The gain then stays constant during hyperinflation and after the hyperinflation is terminated by the ERR. The behavioral interpretation is that the agent realizes that he makes large forecast errors and interprets them as a change in regime, not as an exceptionally large shock in a stationary environment. The agent wants to learn about this new regime as quickly as possible in order not to make large forecast errors, and so he switches to a high constant gain. As soon as forecast errors are below the critical level, the agent resumes the use of a decreasing gain.

To summarize, the model in this paper produces hyperinflation as in Marcet and Nicolini (2003) and, in addition, matches the previous empirical finding that gains increase during hyperinflation.

Another difference in the behavior of the gain occurs after hyperinflation ends. The second panel of Figure 1 illustrates behavior in the MN model, and the second panel of Figure 2 illustrates behavior of the adaptive step-size gain on the left y -axis, and the derivative V on the right y -axis. The agent in this paper *decreases* his gain, whereas the MN agent continues to update at a high constant gain after hyperinflation is ended. The MN agent uses a constant gain because he makes forecast errors above the critical level during hyperinflation and after it. The behavior of the agent using the adaptive step-size algorithm can be described as follows. During hyperinflation, the agent increases his gain to speed up learning, as explained above. By the time hyperinflation reaches its peak, the agent has experienced a long history of underpredicting actual inflation. When hyperinflation is terminated by imposing ERR, the agent makes a negative forecast error ($\pi_t - \beta_t < 0$); i.e., he overpredicts actual inflation. When the forecast error is negative in (8), the agent will reduce his inflation forecast. A decrease in the gain means that the forecast will be reduced by a smaller fraction of the last forecast error and achieve a smaller revision of expectations. This can be interpreted as the agent not wanting to rush to lower his forecast.

There can be several reasons for an agent's unwillingness to update quickly. If the ERR is unknown to the agent, or is perhaps not credible, then the agent will be cautious, and not revise inflation forecasts too quickly. Therefore, he reduces the gain to decrease his response to the negative forecast error. The agent needs to confirm that low inflation is not due to a temporary shock; and if low inflation is the new government's policy, the agent needs to observe that the government is successful in implementing it. After the agent observes low inflation for some time, he will believe it is caused by a change in regime, and thus increase his gain to learn it more quickly. Figure 3 shows that this behavior is typical after the end of hyperinflations.

I would like to provide the technical explanation of why the gain decreases after hyperinflation is stopped. The algorithmic analog of the "long history of underpredicting the inflation" is a high and positive value of V in (9, 10). V stores discounted past forecast errors. By the peak of the hyperinflation, V is high and positive. When hyperinflation ends, negative forecast errors $\pi_t - \beta_t < 0$ lower V according to (10), but not by much, and so V remains positive. From (9), the gain decreases. The key aspect of this behavior is the long history of past mistakes (high positive V) relative to negative forecast error. For the gain to increase, it is necessary to repeatedly experience negative forecast errors such that V decreases and becomes negative. When V and $\pi_t - \beta_t$ are both negative, the gain can increase. This means the agent's response to the latest forecast error increases, i.e., he starts to revise inflation expectations down by larger increments.

This adaptive step-size algorithm produces gain dynamics that are directly related to the learning mechanism specified in Cho and Sargent (1997). Their specification is set up so that the agent is suspicious about the government's implementation of good policy. When applied to a hyperinflationary environment, the model of Cho and Sargent (1997) implies the following behavior. During

TABLE 1. Mean squared error comparisons (standard deviations are in parentheses)

Period	Marcet and Nicolini (2003)	Adaptive step size
100	7.00860 (8.4092)	7.30467 (8.68631)
200	9.21476 (5.46953)	8.71826 (5.48384)
300	11.21859 (5.52852)	8.00149 (4.22205)
400	11.57840 (4.62655)	7.35961 (3.07069)
500	11.74324 (4.29320)	7.10317 (2.80545)

hyperinflation the agent realizes that the government's policy is bad, and so he updates his beliefs with a high gain in order to be alert to possible repercussions. When hyperinflation is ended with the ERR, the agent no longer trusts the government's intentions, or possibly does not believe in the new government's ability to implement the new policy successfully, and so the agent slows down the update of his beliefs.

Next, I evaluate the performance of the forecasting mechanism based on the adaptive step-size algorithm and the mechanism in MN by comparing mean squared errors (MSE) in the simulations. I run 100 simulations of each type, compute MSE for each run, and then average over 100 simulations. The values of model parameters are as described above. I compute MSE for the simulations of different lengths. The results are summarized in Table 1. For a simulation length of 100 periods, the values of MSE are comparable for the two mechanisms. The adaptive step-size gain performs better for simulation lengths greater or equal 200. This means that it takes some time for this specification to be put to its best use.

Next I compare how hyperinflation happens in this model and in other models. In Sargent et al. (2006), hyperinflation occurs when a sequence of seignorage shocks push inflation expectations above the high unstable self-confirming equilibrium (SCE), which means that inflation dynamics escapes the domain of attraction of the low SCE. (SCE are good approximations of rational expectations equilibria for very persistent average deficit states.) In this escape region, actual inflation is higher than expected inflation, and so both actual and perceived inflation increase, and thus hyperinflations occurs. The end of hyperinflation is explained by learning dynamics or changes in fundamentals.

A similar mechanism is in place in Marcet and Nicolini (2003). If inflation starts below the high-inflation steady state, actual inflation is on average closer to the low-inflation steady state than perceived inflation, and so learning moves perceived inflation toward the low-inflation steady state. If perceived inflation is above the high-inflation steady state, then actual inflation is on average higher than perceived inflation, and so perceived inflation increases. There is also an additional amplifying impact of the increased gain during the periods leading to

hyperinflation and during hyperinflation: as the agent makes large forecast errors, the gain increases, and so the agent updates perceived inflation more quickly which feeds back into actual inflation, further increasing it. When inflation reaches the critical level, the ERR is implemented and actual inflation is set to the low-inflation steady state.

Hyperinflation happens in the adaptive step-size algorithm model and in the model of Marcet and Nicolini (2003) in the same way. The difference between Marcet and Nicolini (2003) and the specification in this paper is the behavior of gain during hyperinflation and right after hyperinflation is ended by implementation of the ERR. In MN, the agent updates his forecasts using high gain (0.2) during and after hyperinflation until his forecast error is lower than the critical level. In my model, the agent increases speed of update as hyperinflation develops. The agent meets the ERR reform with suspicion, and lowers the speed of update after he observes a sudden large drop of inflation. After a sufficiently long period of low inflation, the agent is convinced that the reform works and proceeds to update his forecasts quickly to learn the new regime.

4. ESTIMATION OF THE MODEL PARAMETERS

Next, I proceed to estimate the model parameters. This model is stylized, and the data on inflation are nonstationary. These factors can make econometric estimation of the model problematic. The estimation approach in this paper is based on indirect inference [Gourieroux et al. (1993)]. The indirect inference method is useful for complex models with intractable likelihood functions, and the only requirement for estimation is that the model can be simulated. The procedure can be summarized as follows. First, I compute an auxiliary parameter from the actual data on inflation. I use skewness as an auxiliary parameter. Second, I simulate the model for different sets of model parameters, collect data from these simulations, and then compute the skewness of simulated inflation. The objective is to find the set of model parameters for which the distance between the auxiliary parameters from the actual and simulated data is the smallest. In other words, I aim to match the moment (skewness) in actual data and in simulated data.

To match the moments in the data and the model, I use a genetic algorithm—a numerical optimization technique first introduced by Holland (1975) and described in Goldberg (1989), Michalewicz (1996), and Bäck et al. (2000). Among the advantages of using a genetic algorithm are that it starts with a set of random solutions and so does not rely on the starting point, and that it is applicable to discontinuous, nondifferentiable, noisy, multimodal, and other unconventional surfaces [Schwefel (2000)]. Bullard and Duffy (2004) use a simulated method of moments with a genetic algorithm to estimate a growth model with structural breaks. The optimization problem here is to minimize the distance between the auxiliary parameters computed from actual data and the auxiliary parameters computed from simulated data with respect to the values of model parameters.

The nature of the landscape is not known in advance, and so the application of a genetic algorithm is appropriate.

The parameters to be estimated are money demand parameters, γ and ϕ , mean $E(d)$ and standard deviation σ_d of seignorage, and the initial value of the gain a_1 . Parameter $\mu = 0.001$ is fixed because the performance of the adaptive step-size algorithm is not sensitive to this parameter [as shown in Kushner and Yin (2003)].

4.1. Description of the Genetic Algorithm

The algorithm starts with N rules. Each rule consists of the model parameters to be estimated: $(\gamma, \phi, E(d), \sigma_d, a_1)$.

The *initial pool of rules* is generated randomly from uniform distributions with supports that are different for each parameter and are given in Table 2 along with the standard deviations for mutation. Other genetic algorithm parameters are given in Table 3. The ranges for the parameters are chosen taking into consideration the values of the corresponding parameters in the related literature [Sargent and Wallace (1987), Marcet and Sargent (1989b), and Sargent et al. (2006)]. Attention must be paid to the value of mean seignorage because it has the maximum value for which a rational expectation equilibrium exists, and this maximum value is determined by the values of the money demand parameters. The value of mean seignorage is restricted to be close to the maximum value because hyperinflation can arise when mean seignorage is sufficiently close to its maximum value. I deal with these aspects by initializing and performing mutation on the mean seignorage after initialization and mutation of the parameters of the money demand in the following way. For each rule's values of money demand parameters γ, ϕ , I compute the maximum value of mean seignorage as

$$\max(d) = (1 + \gamma - 2\sqrt{\gamma})\phi. \quad (16)$$

Then I restrict the range for mean seignorage $[\min_d, \max_d]$ within some distance from the maximum value $\max(d)$ as $\min_d = 0.9\max(d)$, $\max_d = 0.99\max(d)$.

To evaluate the performance of each rule, the simulation is run for the parameter values of this rule, simulated inflation data are collected, and skewness of the simulated inflation is computed. The length of the simulation is equal to the number of observations in the data on inflation plus 20 periods. The initial 20 periods of simulated data are discarded to avoid dependence on the initial values.

The fitness criterion is very important for the performance of the algorithm. Mean squared error is often used as a performance measure. In this estimation, fitness is the squared distance between skewness in the actual data and skewness in the simulated data computed and added for 100 realizations.

Once fitness is computed, the pool of rules is updated by replication, crossover, and mutation.

Replication is done by tournament selection. Two rules are randomly selected with replacement from an old pool of rules. The fitness measures of these two

TABLE 2. Ranges and standard deviations for mutation for different parameters

Parameter	Range	Standard deviation
γ	[0.3, 0.9]	0.1
ϕ	[0.3, 1.9]	0.1
$E(d)$	Depends on γ, ϕ	0.05
σ_d	[0.01, 0.2]	0.05
a_1	[0.01, 0.4]	0.05

TABLE 3. Genetic algorithm parameters

Parameter	Value
N	30
T	500
pcross	0.5
pmut	0.1

rules are compared, and the rule with higher fitness is chosen into the new pool of rules. This procedure is repeated N times to form a new pool of rules. Replication provides all the selection pressure in this genetic algorithm.

Crossover is done with probability $\text{pcross} = 0.5$ for a pair of randomly selected (without replacement) rules. Once a pair of rules is chosen for crossover, the values of each parameter are exchanged between the rules with probability 0.5.

TABLE 4. Estimated values of the model parameters

Country sample	Skewness [data]	γ	ϕ	$E(d)$	σ_d	a_1
Argentina	6.2538	0.35	1.17	0.194	0.18	0.05
1984:01–2000:04	[5.6072]	(0.11)	(0.11)	(0.06)	(0.01)	(0.045)
Bolivia	6.2021	0.38	0.71	0.099	0.16	0.05
1982:02–2001:09	[5.2451]	(0.10)	(0.13)	(0.03)	(0.02)	(0.038)
Brazil	1.1931	0.34	1.82	0.068	0.10	0.04
1984:01–2004:09	[1.9263]	(0.11)	(0.31)	(0.13)	(0.04)	(0.10)
Peru	8.2743	0.77	1.62	0.019	0.17	0.07
1988:04–2004:09	[10.5472]	(0.09)	(0.23)	(0.002)	(0.02)	(0.07)

Notes: The standard deviations of the estimated parameters are given in the parenthesis. The first row of Column (2) gives skewness of the simulated data based on the estimated parameters, the second row of Column (2) gives skewness in the data.

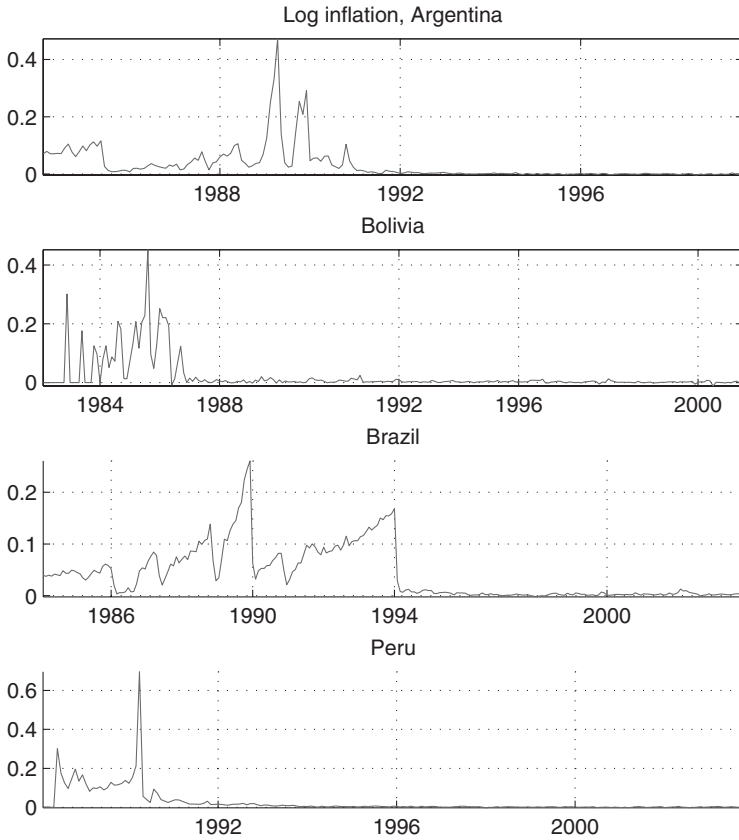


FIGURE 4. Monthly inflation rates (in logs).

Mutation is done for each parameter in the rule with probability $p_{mut} = 0.1$. The new value of the parameter is computed as:

$$\text{new rule} = \text{old rule} + \text{randn} \times \text{std} \tag{17}$$

where randn is the random number from the standard normal distribution $(N(0,1))$, and std is the standard deviation for a specific parameter. The values of std for each parameter are presented in Table 2. If the value of a new rule is below (above) the low (high) boundary value, the new rule is set equal to the low (high) boundary value.

The genetic algorithm is repeated for T iterations. The purpose is for the parameters to converge to their globally optimal values by the end of the genetic algorithm simulation. The genetic algorithm parameters are presented in Table 3.

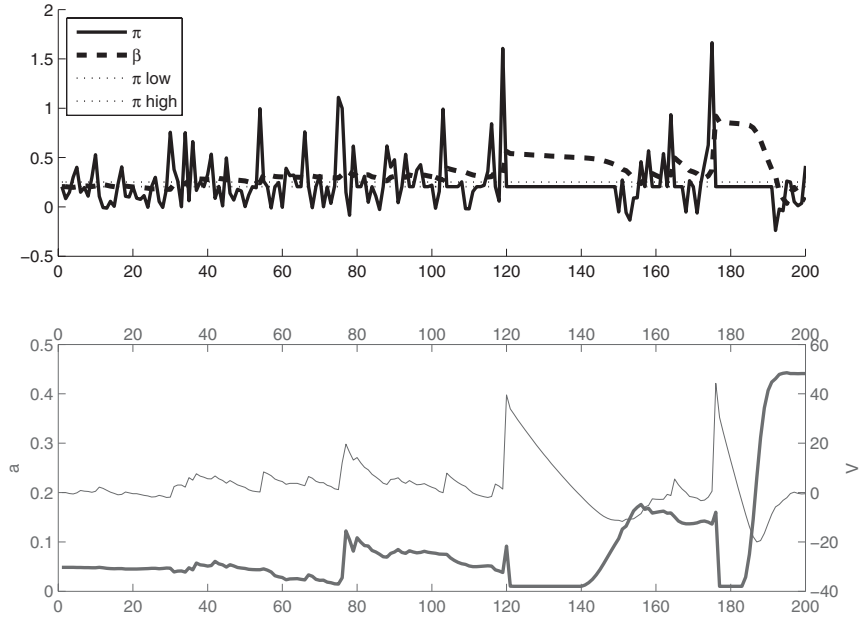


FIGURE 5. Typical simulation with estimated parameters based on skewness for Argentina. The first panel shows actual inflation, π , and expectations of inflation, β . The second panel shows adaptive step-size gain on the left y-axis (thick line) and the derivative V on the right y-axis (thin line).

4.2. Results of the Computations

I now present the estimated parameters from the genetic algorithm optimization, with skewness as an auxiliary parameter. For each parameter, I report the average and standard deviation (in parenthesis) of thirty rules after 40,000 iterations of the genetic algorithm.

The estimated parameter values for Argentina, Bolivia, Brazil and Peru are presented in Table 4. The data used for these countries are monthly inflation rates for the samples indicated in Table 4 and computed from International Financial Statistics Consumer Price Indexes. The data are plotted in Figure 4. The skewness in data for each country is given in Column (2) in parentheses. The skewness for simulated data for the estimated parameters is given in Column (2).

Using these estimated parameter values, I simulate the model to obtain the path of the time-varying gain and present the simulations in Figure 5 for Argentina and in Figure 6 for Bolivia. For all countries except Brazil, the model with estimated parameters exhibits hyperinflation. The figures of the simulated data show that the model captures the general features in the actual data when compared to the figure with the actual data shown in Figure 4. The second panel of Figures 5 and 6 shows that the gain increases during hyperinflation and falls right after hyperinflation

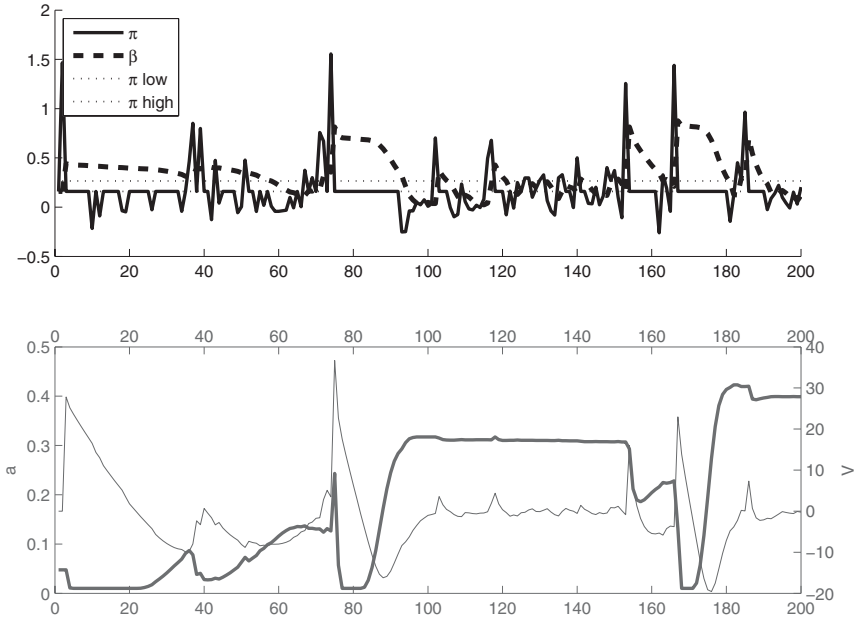


FIGURE 6. Typical simulation with estimated parameters based on skewness for Bolivia. The first panel shows actual inflation, π , and expectations of inflation, β . The second panel shows adaptive step-size gain on the left y-axis (thick line) and the derivative V on the right y-axis (thin line).

ends. Capturing this behavior is the distinguishing feature of the model with the adaptive step-size algorithm.

Matching model skewness with skewness in the data can be viewed as a first step. A more sophisticated approach may be warranted, for example, matching transition probabilities in inflation data and simulated inflation for an autoregressive Markov-switching model.

5. CONCLUSION

This paper presents the first application of an adaptive step-size algorithm to model time-varying learning. The performance of the algorithm is illustrated in the environment of Marcet and Nicolini (2003). The model behaves qualitatively similarly to that of Marcet and Nicolini (2003) and performs quantitatively somewhat better, based on the criterion of mean squared error. The agent increases the speed of update during hyperinflation and lowers it after the hyperinflation ends. This behavior matches the findings in the data. The agent using this model shows caution when faced with sudden changes in policy, and is able to recognize the change in regime after acquiring sufficient information. The performance of this new algorithm in models of hyperinflation suggests that it might be useful in less extreme environments as well, e.g., in conventional models of asset pricing and

business cycles. One could also follow Evans and Ramey (2006) and extend their notion of an equilibrium gain parameter to a setting with a time-varying gain.

NOTES

1. The literature discussed in this paragraph often uses the term “coefficient of expectations” for the gain or update speed.
2. I would like to thank the referee for providing this interpretation.
3. For a constant gain a , equation (10) for V_t can be written as $V_t = \sum_{k=0}^{t-1} (1-a)^k (\pi_t - \beta_t)$, and it is a weighted average of past forecast errors.
4. The model in Sargent et al. (2006) is estimated with seignorage as a Markov switching process, and so their values of the model parameters are different from those of Marcet and Nicolini (2003).
5. This is the value of constant gain that Marcet and Nicolini (2003) use.
6. This is because the initial prediction error is zero, and so it is below a threshold that causes a switch into constant gain in the expectation scheme of Marcet and Nicolini (2003).

REFERENCES

- Adam, Klaus, George W. Evans, and Seppo Honkapohja (2006) Are hyperinflation paths learnable? *Journal of Economic Dynamics and Control* 30, 2725–2748.
- Bäck, Thomas, David Fogel, and Zbigniew Michalewicz (2000) *Evolutionary Computation 1. Basic Algorithms and Operators*. Bristol and Philadelphia: Institute of Physics Publishing.
- Benveniste, Albert, Michel Metivier, and Pierre Priouret (1990) *Adaptive Algorithms and Stochastic Approximations*. New York: Springer-Verlag.
- Branch, William A. and George W. Evans (2006) A simple recursive forecasting model. *Economics Letters* 91, 158–166.
- Bullard, James B. (1991) Learning, rational expectations and policy: A summary of recent research. *Federal Reserve Bank of St. Louis Review* January–February, 50–60.
- Bullard, James B. (2006) The learnability criterion and monetary policy. *Federal Reserve Bank of St. Louis Review* 88(3), 203–217.
- Bullard, James B. and John Duffy (2004) Learning and Structural Change in Macroeconomic Data. Federal Reserve Bank of St. Louis working paper 2004–016A.
- Bullard, James B. and Kaushik Mitra (2002) Learning about monetary policy rules. *Journal of Monetary Economics* 49(6), 1105–1129.
- Cagan, Phillip (1956) The monetary dynamics of hyperinflation. In Milton Friedman (ed.), *Studies in the Quantity Theory of Money*, pp. 25–117. Chicago: The University of Chicago Press.
- Chakraborty, Avik and George W. Evans (2008) Can perpetual learning explain the forward premium puzzle? *Journal of Monetary Economics* 55, 477–490.
- Cho, In-Koo and Kenneth Kasa (2008) Learning dynamics and endogenous currency crisis. *Macroeconomic Dynamics* 12(2), 257–285.
- Cho, In-Koo and Thomas J. Sargent (1997) Learning to Be Credible. Working paper. Available at <http://homepages.nyu.edu/~ts43/RESEARCH/research.htm#Learning>.
- Evans, George W. and Seppo Honkapohja (1993) Adaptive forecasts, hysteresis, and endogenous fluctuations. *Federal Reserve Bank of San Francisco Economic Review* 1, 3–13.
- Evans, George W. and Seppo Honkapohja (2001) *Expectations and Learning in Macroeconomics*. Princeton, NJ: Princeton University Press.
- Evans, George W. and Garey Ramey (2006) Adaptive expectations, underparameterization and the Lucas critique. *Journal of Monetary Economics* 53, 249–264.
- Funke, Michael, Stephen Hall, and Martin Sola (1994) Rational bubbles during Polands’ hyperinflation: Implications and empirical evidence. *European Economic Review* 38 (6), 1257–76.
- Goldberg, David E. (1989) *Genetic Algorithms in Search, Optimizations, and Machine Learning*. Reading, MA: Addison–Wesley.

- Gourieroux, Christian, Alain Monfort, and Eric Renault (1993) Indirect inference. *Journal of Applied Econometrics*. Issue Supplement: Special Issue on Econometric Inference Using Simulation Techniques 8, 85–118.
- Holland, John H. (1975) *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.
- Khan, Mohsin S. (1977) The variability of expectations in hyperinflations. *Journal of Political Economy* 85 (4), 817–827.
- Kushner, Harold J. and Jichuan Yang (1995) Analysis of adaptive step-size SA algorithms for parameter tracking. *IEEE Transactions on Automatic Control* 40 (8), 1403–1410.
- Kushner, Harold J. and G. George Yin (2003) *Stochastic Approximation and Recursive Algorithms and Applications*. 2nd ed. New York: Springer-Verlag.
- Kydland, Finn E. and Edward C. Prescott (1977) Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy* 85, 473–491.
- Ljung, Lennart and Torsten Söderström (1983) *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press.
- Marcet, Albert and Juan Pablo Nicolini (2003) Recurrent hyperinflations and learning. *American Economic Review* 93(5), 1476–1498.
- Marcet, Albert and Thomas J. Sargent (1989a) Convergence of least squares learning mechanisms in self-referential linear stochastic models. *Journal of Economic Theory* 48, 337–368.
- Marcet, Albert and Thomas J. Sargent (1989b) Least squares learning and the dynamics of hyperinflations. In William A. Barnett, John Geweke, and Karl Shell (eds.), *Economic Complexity: Chaos, Sunspots, bubbles, and Nonlinearity: Proceedings of the Fourth International Symposium in Economic Theory and Econometrics*, pp. 119–137. Cambridge, UK: Cambridge University Press.
- Michalewicz, Zbigniew (1996) *Genetic Algorithms + Data Structures = Evolution Programs*, 3rd ed. Berlin: Springer-Verlag.
- Milani, Fabio (2007) Expectations, learning and macroeconomic persistence. *Journal of Monetary Economics* 54(7), 2065–2082.
- Milani, Fabio (2008) Learning, monetary policy rules, and macroeconomic stability. *Journal of Economic Dynamics and Control* 32(10), 3148–3165.
- Mussa, Michael (1981) Sticky prices and disequilibrium adjustment in a rational model of the inflationary process. *American Economic Review* 71(5), 1020–1027.
- Orphanides, Athanasios and John C. Williams (2004) The Decline of Activist Stabilization Policy. Board of Governors of the Federal Reserve System working paper 804.
- Orphanides, Athanasios and John C. Williams (2005) Imperfect knowledge, inflation expectations, and monetary policy. In Ben S. Bernanke and Michael Woodford (eds.), *The Inflation-Targeting Debate*, pp. 201–234. Chicago: NBER and University of Chicago Press.
- Sargent, Thomas J. (1999) *The Conquest of American Inflation*. Princeton, NJ: Princeton University Press.
- Sargent, Thomas J. and Neil Wallace (1987) Inflation and government budget constraint. In Assaf Razin and Efraim Sadka (eds.), *Economic Policy in Theory and Practice*, pp. 170–200. New York: St. Martin's Press.
- Sargent, Thomas J., Noah Williams, and Tao A. Zha (2006) The Conquest of South American Inflation. NBER working paper series 12606.
- Schwefel, Hans-Paul (2000) Advantages (and disadvantages) of evolutionary computation over other approaches. In Thomas Bäck, David Fogel, and Zbigniew Michalewicz (eds.), *Evolutionary Computation I. Basis Algorithms and Operators*. Bristol and Philadelphia: Institute of Physics Publishing.
- Silveira, Antonio M. (1973) The demand for money: The evidence from the Brazilian economy. *Journal of Money, Credit and Banking* 5(1), 113–140.
- Timmermann, Allan G. (1993) How learning in financial markets generates excess volatility and predictability is stock prices. *Quarterly Journal of Economics* 108 (4), 1135–1145.