Macroeconomic Dynamics, **12**, 2008, 257–285. Printed in the United States of America.
DOI: 10.1017/S1365100507070046

LEARNING DYNAMICS AND ENDOGENOUS CURRENCY CRISES

IN-KOO CHO
University of Illinois

KENNETH KASA
Simon Fraser University

This paper introduces adaptive learning into the third-generation currency crisis model of Aghion, Bacchetta, and Banerjee (2001, Currency crises and monetary policy in an economy with credit constraints, *European Economic Review* 45, 1121–1150). Adaptive learning might reflect, for example, uncertainty about the economy's exposure to adverse balance sheet effects. Even when equilibrium is unique, we show that the learning algorithm's escape dynamics can produce the same kind of Markov-switching exchange rate behavior that is typically attributed to sunspots or herds. An advantage of our learning model is that currency crises become endogenous, in the sense that their stochastic properties can be related to assumptions about learning and other structural features of the economy.

Keywords: Learning, Currency crises

1. INTRODUCTION

Economists have made great strides during the past decade in understanding currency crises. Following the ERM Crisis (1992–1993), Obstfeld (1994) developed a class of models based on an open-economy version of the Barro–Gordon model, which explained many of the puzzling features of this episode. Contrary to the predictions of the prevailing first-generation models, countries that left the EMS or widened their intervention bands did not do so because they "ran out of reserves." Instead, the decisions seemed to be motivated by the desire to avoid the unpleasant macroeconomic consequences of remaining in the system. Obstfeld's model formalized these trade-offs and offered new insights into the nature of currency crises. He showed that when governments choose exchange rates sequentially in order to minimize a loss function, currency crises can become self-fulfilling prophecies, in the sense that expectations of a devaluation can elicit (ex post optimal)

We would like to thank Phillipe Bacchetta, Fabrizio Perri, Laura Veldkamp, and participants at the AEA, SED, and Econometric Society meetings for helpful comments on an earlier version. We owe special thanks to two anonymous referees. Financial support from the National Science Foundation (SES-0004315, ECS-0523620) is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Address correspondence to: Kenneth Kasa, Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada; e-mail: kkasa@sfu.ca.

1365-1005/08 $18.00

responses by the government that ratify those beliefs. Obstfeld's work triggered a flood of research during the 1990s on multiple equilibria in foreign exchange markets.[1]

Unfortunately, these so-called second-generation models encountered empirical problems almost immediately. One of the leading stylized facts of the Mexican and Asian crises was the combination of devaluation and subsequent recession. Instead of devaluing in order to *avoid* a recession, the devaluations of Mexico and Asia seemed to be *causing* a recession. Clearly, something was missing from second-generation currency crisis models.

Although there are many reasons that a devaluation might prove to be contractionary, the recent literature on third-generation currency crisis models has focused on the role of foreign currency–denominated debt and its adverse "balance sheet effects."[2] This focus is empirically motivated, because foreign currency debt seemed to be at the heart of both the Mexican and Asian crises.[3] In Mexico's case it was primarily the government that was exposed, whereas in Asia it was primarily the private sector. Either way, a sudden devaluation erodes net worth, and to the extent that investment and borrowing capacity is constrained by net worth, due perhaps to information and incentive problems, expectations of a devaluation can turn out to be just as self-fulfilling as in the earlier second-generation class of models. The crucial difference is that now devaluations produce recessions.

Notwithstanding their contrasting predictions about the output effects of devaluations, second- and third-generation currency crisis models share one important feature—both models interpret a crisis as a sudden switch to a "bad equilibrium." As is now well known, when a model exhibits multiple equilibria, it is usually possible to layer on an exogenous sunspot process that governs switches between them. It is this exogenous sunspot process that is ultimately to blame for currency crises in these models.[4]

A more recent class of models that do link crises endogenously to fundamentals view these events as "herds" or informational "cascades" [e.g., Chari and Kehoe (2004)]. Like sunspot models, these models have the attractive property that crises are tenuously linked to macroeconomic fundamentals and seem to come out of nowhere. Unlike sunspot models, however, they fully account for the underlying decisions that generate a crisis. Unfortunately, herding models do not explain a key observed feature of crises, namely, their *recurrence*. Most countries that experience financial crises do so more than once. They have a *history* of instability. In existing herding models, either a crisis happens or it does not. If it does, the game is over. There is no linkage between the past occurrence of crises and the likelihood of future crises.

Our paper tries to explain both why crises erupt suddenly, often without macroeconomic warning, and their recurrence. We do this by abandoning a key assumption of both the sunspot literature and the herding literature, namely, the Rational Expectations Hypothesis. Instead, we assume that agents must form their beliefs adaptively, without a priori knowledge of the economy's underlying structure. In a sense, we share the same misgivings as Morris and Shin (1998), who express doubts about the common knowledge assumptions of existing currency crisis models. Like theirs, our approach eliminates multiple equilibria and avoids unpleasant questions about how agents suddenly coordinate their expectations on exogenous sunspots. A key result of our paper is to show that despite the fact that equilibrium is unique in our model, this equilibrium is a *stochastic process*, and as a stochastic process it features exactly the sort of Markov-switching dynamics that is usually attributed to sunspots. An advantage of our approach is that the stochastic properties of crises can be related to assumptions about learning and other structural features of the economy. In contrast, sunspot models place no testable restrictions on the dynamics of currency crises. They merely rationalize their occurrence ex post.[5] Also, in contrast to herding models, our model accounts for the crucial role of history in cultivating the conditions that are conducive to currency crises.

Of course, we do pay a price for abandoning the Rational Expectations Hypothesis, because we are forced into the "wilderness of bounded rationality." As noted by Sargent (1993), bounded rationality is a wilderness because, whereas there is only one way to get it right, there are an infinite number of ways to be wrong. In this paper, we consider an extension of the well-known least-squares learning approach, as surveyed by Evans and Honkapohja (2001). Agents are assumed to specify linear regression models and recursively estimate their parameters. This approach introduces bounded rationality because it assumes that agents fail to recognize and respond to their own influence over future data. They learn, but in a purely passive, retrospective way. In contrast, a Bayesian would recognize that the data-generating process is evolving through time and exploit this variation in his inferences and policies.[6]

We extend this traditional learning approach following the pioneering work of Sargent (1999). Rather than postulate a correctly specified model, and then ask whether agents eventually zero in on the correct parameter values, Sargent imputes a subtle form of specification error to the government. Whereas in reality the economy has a natural rate structure, in the sense that only unanticipated policy actions matter, the government mistakenly believes in a Keynesian model, where the systematic component of policy matters.[7] As a result, the evolving beliefs of the private sector inject "parameter drift" into the government's model, which the government responds to by placing more emphasis on recently observed data. This is accomplished by the use of a constant-gain stochastic approximation algorithm.

With constant-gain learning, beliefs do not converge to a fixed limit. Instead, they converge (weakly) to a stochastic process. Sargent (1999) and Cho et al. (2002) (henceforth denoted CWS) show that this stochastic process features recurrent nonlinear dynamics, which resembles a Markov-switching process, and which can be characterized using the tools of large deviations theory. Our central contention is that this nonlinear "escape dynamics" could be a contributing factor in observed currency crises. They occur as the government vacillates between

Sargent's two observationally equivalent ways of interpreting the data. In our particular model, this happens when the government confuses the natural rate structure of the economy with the apparent absence of balance sheet effects.

Although surely not the whole story, self-referential learning dynamics do offer a fresh perspective on the recent currency crisis literature. Debates about the causes and consequences of currency crises often focus on the distinction between "bad luck" and "bad policy," i.e., between sunspots and fundamentals. The policy implications of this distinction are important. If crises are the result of bad luck, then one can argue that IMF-style bailout policies make sense. On the other hand, if crises result from bad policy, then bailouts are much harder to defend. Interestingly, our analysis suggests that it is the *interaction* of bad luck and bad policy that is ultimately to blame. Good policy can overcome bad luck, and good luck can sustain a bad policy. We show, however, that when governments misinterpret the instability of their models in a particular, and we believe plausible, way, then rare but recurring sequences of shocks can suddenly trigger what may appear to be sunspot-induced, self-fulfilling currency crises. Moreover, large deviations methods provide a precise analytical characterization of these shocks. They also provide estimates of crisis frequency as a function of the economy's underlying parameters. These results represent a significant advance over sunspot and herding models, in the sense that they place *testable restrictions* on the data.

The remainder of the paper is organized as follows. Section 2 develops our baseline third-generation crisis model. Although there are many possible models we could use as a platform, we employ a version of the model of Aghion et al. (2001; hereafter denoted ABB). We use this model because of its familiarity to many readers and its simplicity.

As a prelude to our analysis of learning dynamics, Section 3 first characterizes Nash and Ramsey equilibria under rational expectations. Although the Nash equilibrium features some endogenous exchange rate fluctuations, due to a time-varying incentive to revalue the currency, these fluctuations are smooth and do not resemble currency crises.

Section 4 discusses constant-gain recursive learning. We begin by defining the notion of a *self-confirming equilibrium* (SCE). An SCE can be interpreted as a weakening of a rational expectations equilibrium, which allows for model misspecification. We show that under reasonable parameter conditions the ABB model has a unique E-stable SCE. The recursive learning dynamics consist of two parts: (1) the *mean dynamics* and (2) the *escape dynamics*. The mean dynamics reflects the efforts of agents to eliminate systematic forecast errors. These push the system toward the SCE. The escape dynamics is driven by rare but recurrent shocks that push the system away from the SCE. The escape dynamics is what drives currency crises in our model. It can be characterized using large deviations methods. Large deviations theory can be thought of as a refinement of the central limit theorem, which permits a rigorous analysis of rare events.

Section 5 relates the analysis of our paper to the existing literature. In a sense, our paper stands the logic of first-generation models on its head. First-generation models view government policy as exogenous and focus on speculation by the private sector. Our paper takes private sector actions as exogenous, in the sense that foreign currency debt is given. Instead, we focus on the government's efforts to cope with model uncertainty, specifically its efforts to learn about the economy's balance sheet exposure over time, which are complicated by the fact that its own actions influence the information content of the data it observes. Surely, the truth lies in the middle. Finally, Section 6 contains some concluding remarks, and an Appendix contains proofs of some technical results.

2. A THIRD-GENERATION CURRENCY CRISIS MODEL

In this section we outline the model of Aghion et al. (2001). Because we are primarily interested in learning dynamics, the presentation here will be brief. The reader should consult ABB's paper for full details.

The defining characteristic of third-generation crisis models is the presence of unhedged foreign currency liabilities (i.e., balance sheet effects), which make (unanticipated) devaluations contractionary. Particular models differ according to who incurs the liabilities and why. For example, in the models of Burnside et al. (2001) and Dooley (2000), it is the banking sector that is exposed, whereas in the models of ABB, Cespedes et al. (2004), and Krugman (1999), it is firms that are exposed. Generally speaking, models that focus on the exposure of the banking sector tend to attribute the exposure to government deposit guarantees, whereas models that focus directly on firms tend to blame the exposure on asymmetric information problems. We follow ABB and attribute balance sheet effects to moral hazard.

Most of the attention in this literature focuses on the combination of these balance sheet effects with financial market imperfections, which cause borrowing to be constrained by net worth. As the literature has demonstrated, this combination creates a potent propagation mechanism. With net foreign currency liabilities, devaluations erode net worth. Then, if borrowing is constrained by net worth, the decline in net worth produces a decline in investment and output, which then reinforces the original exchange rate decline. As in second-generation models, this circularity exposes the economy to multiple equilibria and sunspot fluctuations.

The ABB model combines three essential ingredients. First, prices are assumed to be preset one period in advance. This produces real effects from nominal exchange rate changes. Second, financial market imperfections limit borrowing to be an endogenously determined multiple of net worth. This creates a "financial accelerator." Third, firms are assumed to be financed, at least partially, by foreign currency debt. As a result, exchange rate changes trigger the financial accelerator.

In full generality, these assumptions would produce a model that is quite complex. The contribution of ABB is to come up with a tractable formulation. We now proceed to outline this formulation.

2.1. Production and Price-Setting

Output of a single good is produced by competitive consumer/entrepreneurs according to a linear production function,

$$Y_t = Ak_t\varepsilon_t, \tag{1}$$

where Y_t is output, k_t is the capital stock, and ε_t is a (positive) productivity shock. ABB assume that capital completely depreciates between periods. In contrast, we assume that capital does not depreciate.

Following ABB, we assume that domestic entrepreneurs face a competitive fringe of foreign producers. Foreign firms have constant marginal costs. Both domestic and foreign firms must set prices at the beginning of each period, before the realization of the production shock and the exchange rate. Assuming foreign marginal costs are constant, and normalizing them to unity, implies that domestic firms must then set $P_t = E_{t-1}S_t$, where S_t denotes the nominal exchange rate, defined as the price of foreign currency. Hence, PPP holds ex ante, but not necessarily ex post.

2.2. The Credit Multiplier and the Currency Composition of Debt

Capital consists of the entrepreneur's own wealth, w_t, and any additional borrowed funds, d_t. That is, $k_t = w_t + d_t$. Firms can borrow either in terms of domestic currency, at interest rate i_{t-1}, or in terms of foreign currency, at (constant) interest rate i^*. As with price-setting, investment decisions must be made at the beginning of the period (so that the loan rate is the prevailing, prior period rate).

Debt contracts are only partially enforceable. In particular, borrowers can pay a cost, cP_tk_t, proportional to the amount invested, that allows them to abscond with the funds. However, if a borrower does default, there is a probability, p, that the lender is able to track him down and collect anyway. Hence, assuming for now a domestic currency loan, a borrower will choose to repay if and only if

$$P_tY_t - (1 + i_{t-1})P_{t-1}d_t \geq P_tY_t - cP_tk_t - p(1 + i_{t-1})P_{t-1}d_t.$$

Collecting terms gives us the incentive compatibility constraint, $d_t \leq \mu_t w_t$, where the "credit multiplier," μ_t, is given by

$$\mu_t = \frac{c}{(1-p)(1+r_{t-1}) - c}, \tag{2}$$

where r_{t-1} is the real interest rate.

There are several things to note about this multiplier. First, it increases with p. That is, firms can borrow more when the "monitoring technology" improves. ABB interpret this as a proxy for financial market development. Second, because lending decisions are made before any shocks are realized, μ_t will be independent of the currency denomination of debt as long as uncovered interest parity and (ex ante) PPP hold. Third, notice that μ_t is state-dependent. That is, it varies with the real interest rate. Later, when we incorporate learning, we will approximate this dependence. Finally, notice that this model of debt is quite different from the influential model of Kehoe and Levine (1993). They assume that debt contracts are enforced by future exclusion from the capital market. In contrast, borrowers in this model start each period with a clean slate.

As noted earlier, firms are free to borrow in either currency. One of the critical questions to emerge in the wake of recent financial crises is why a domestic economy would choose to become so exposed. That is, why is there so much unhedged foreign currency borrowing? By now, there are many (not necessarily mutually exclusive) theories. Perhaps the most common explanation relies on government bailout guarantees.[8] Alternatively, Jeanne (2000) shows that foreign debt might play a signaling role, which lowers its interest rate. Tirole (2003) argues that foreign currency debt can provide a commitment device to domestic governments wanting to attract capital inflows. Another possibility is to appeal to moral hazard, as in ABB. They show that if the currency composition of a borrower's debt is not observable, it might be optimal to borrow abroad. All of these theories imply a distinction between privately optimal and socially optimal financing decisions, because individual firms ignore the effects of their borrowing decisions on the country's financial fragility.

Which of these theories is more important is irrelevant for us, because we follow ABB and assume that foreign debt is exogenous. In particular, we assume that the foreign currency value of foreign debt is held constant at $\bar{d}^*$. This is a tremendous simplification. It might not be such a bad assumption in models such of Aghion et al. (2004) and Burnside et al. (2001), where firms are led to a corner solution, and foreign debt is at its (credit-constrained) maximum. Later we discuss how our results might be altered if the level and composition of debt were endogenous.

In general, the dynamics of wealth and output depend on whether the borrowing constraint is binding. In fact, ABB emphasize that the distinction between a binding and a nonbinding borrowing constraint lies at the heart of the model's nonlinearity and its capacity to generate multiple equilibria. However, because generating multiple equilibria is not our goal, we further assume that the borrowing constraint binds in every period. This implies restrictions on the production function and the support of the shocks. For example, it must be the case that $A\mathsf{E}(\varepsilon) > (1 + i^*)(S^{\mathrm{e}}/S)$ for all values of S^{e}/S, where S^{e}/S is the expected (percentage) depreciation of the domestic currency. Otherwise, the firm would rather invest abroad.

When the borrowing constraint is binding, output can be written as a function of the entrepreneur's wealth:

$$Y_t = A(1 + \mu_t) w_t \varepsilon_t. \tag{3}$$

Note, however, that because we are now treating foreign debt exogenously, the proportionality between output and wealth only applies if foreign borrowing is not (directly) productive. It still plays an important indirect role, however, because unexpected capital gains and losses on foreign debt affect profits and wealth, which influence domestic borrowing capacity.

To derive the law of motion for output we therefore need to derive the law of motion for wealth. Following ABB (2001), we assume that entrepreneurs consume (if they can) a fixed fraction, χ, of their wealth. If wealth is zero they consume nothing. Without depreciation, end-of-period wealth therefore evolves according to

$$w_t = (1-\chi)\left[w_{t-1} + \frac{\Pi_t}{P_t}\right], \tag{4}$$

where Π_t represents (nominal) profits net of debt repayments,

$$\Pi_t = P_tY_t - (1+i_{t-1})P_{t-1}d_t - (1+i^*)\frac{S_t}{S_{t-1}}P_{t-1}d_t^*, \tag{5}$$

where d_t is the real value of domestic debt and d_t^* is the real value of foreign debt. The timing is as follows: (1) Entrepreneurs borrow funds at the beginning of the period, at the last period's end-of-period interest rate; (2) shocks are realized and observed; and then (3) domestic debt repayment decisions are made. Notice that in contrast to ABB, entrepreneurs can effectively use current borrowing to support current production and more current borrowing. That is, having w_t depend on Π_t creates a kind of simultaneity, in the sense that d_t depends on w_t, which in turn depends on d_t. Of course, for this to make sense, there must be multiple lenders, who cannot observe what others have lent. It also means that for a well-defined equilibrium to exist, certain parameter restrictions (discussed below) must be satisfied.

Combining (3)–(5), and using the assumption that nominal foreign currency debt is constant (i.e., $P_{t-1}d_t^*/S_{t-1} = \bar{d}^*$), as well as the assumption that the domestic borrowing constraint always binds, which permits Y_{t-1} to be substituted for w_{t-1} and Y_t to be substituted for d_t, delivers the following law of motion for output:

$$\begin{aligned} Y_t = {} & \frac{1}{1-(1-\chi)[A\varepsilon_t + \mu_t(A\varepsilon_t - (1-r_{t-1}))]} \\ & \times \left\{ \frac{\varepsilon_t(1+\mu_t)}{\varepsilon_{t-1}(1+\mu_{t-1})}(1-\chi)Y_{t-1} - A\varepsilon_t(1+\mu_t)(1-\chi)(1+i^*)\bar{d}^*\frac{S_t}{P_t} \right\}. \end{aligned} \tag{6}$$

For an equilibrium to exist it must be the case that $(1-\chi)[A\varepsilon_t + \mu_t(A\varepsilon_t - (1+r_{t-1}))] \neq 1 \; \forall t$ and all realizations. In addition, stationarity requires that the coefficient on Y_{t-1} be less than one sufficiently often. These restrictions can be interpreted as placing an upper bound on A (as a function of χ and the shocks). At the same time, however, remember that a binding borrowing constraint places a *lower* bound on A. We assume that these bounds can both be satisfied.

Equation (6) is one of the two key equations of the model. Notice the role of balance sheet effects. Because prices are predetermined, a depreciation raises the foreign debt burden, which exerts a contractionary effect on output. However, this is not quite the end of the story. As noted by ABB, because μ_t depends negatively on r_{t-1}, and since r_{t-1} depends negatively on S_{t-1}/P_{t-1} [due to uncovered interest parity, $1 + r_t = (1 + i^*)(P_t/S_t)$], unanticipated depreciations also relax the borrowing constraint for a given level of wealth, since they reduce domestic real interest rates. This exerts an offsetting expansionary effect on output. So in general, as you might expect, the effect of a surprise depreciation on output is ambiguous. What is clear is that the negative effect is more likely to dominate when $\bar{d}^*$ is larger.

As it stands, (6) is still too complicated to be useful, despite the many assumptions that have already been made in deriving it. Therefore, we make one last assumption and take a log-linear approximation. Because expectations are the focus of our analysis, when doing this we use the pricing rule to substitute $E_{t-1}S_t$ in place of P_t. This gives us

$$y_t = (1 - \rho)\bar{y} + \rho y_{t-1} + \alpha(s_t - E_{t-1}s_t) + \sigma_1 v_{1t}, \tag{7}$$

where lowercase letters are natural logs of the corresponding uppercase letters, and where the $N(0, 1)$ error term, v_{1t}, combines the productivity shock ε_t and approximation errors. It is presumed to be i.i.d.

Equation (7) is a standard open-economy "expectations-augmented Phillips Curve," with one crucial exception. In this model the slope of the Phillips Curve is indeterminate. As noted earlier, the sign of α depends on the relative importance of balance sheet effects. In what follows, we assume that balance sheet effects are relatively strong, so that $\alpha < 0$.

2.3. Financial Markets and Monetary Policy

ABB (2001) close their model by combining the uncovered interest parity condition with a standard money demand equation. This delivers a second equation relating the exchange rate to future output, called the IPLM curve. Changes in the money supply shift the IPLM curve. This allows ABB to make statements about how monetary policy should respond to the fait accompli of a currency crisis.

In this paper, we assume the government chooses the exchange rate to minimize an explicit intertemporal loss function. This loss function reflects a trade-off between output stability and exchange rate stability,

$$\min_{\{s_t\}} E_t \frac{1}{2} \sum_{j=0}^{\infty} \beta^j \left[(s_{t+j} - s^*)^2 + \lambda(y_{t+j} - y^*)^2\right], \tag{8}$$

where the parameters s^* and y^* are arbitrary targets, and λ captures the relative cost of output fluctuations. A fixed exchange rate regime, albeit an uninteresting one, would result if $\lambda = 0$.

Of course, this is an undeniably ad hoc objective function. Although this prevents us from making serious welfare statements, we nonetheless view it as reasonable from a descriptive standpoint. In fact, something closely resembling this objective can be derived as a quadratic approximation to a utility-based welfare criterion, although this would likely require a separate treatment of exchange rate and price level instability [see, e.g., Clarida et al. (2001)].

Although in practice few central banks think of themselves as setting the exchange rate directly, absent an explicit and distinct cost of interest rate volatility there is little loss of generality in assuming that the central bank sets the exchange rate rather than the interest rate. Given an exchange rate policy, we can use the following (risk-adjusted) uncovered interest parity condition to infer the model's implied interest rate path:

$$i_t = i^* + E_t s_{t+1} - s_t - \phi(y_{t-1} - y^{ss}), \tag{9}$$

where the ϕ parameter captures the effect of net worth on the risk premium.[9]

3. RATIONAL EXPECTATIONS EQUILIBRIA

As ABB readily acknowledge, they do not discuss the potential importance of expectations and credibility in their model. Instead, they confine their attention to purely temporary, totally unanticipated shocks. From inspection of (7) and (9), however, it is clear that these issues are going to be central in any real world setting. Devaluations that are anticipated will be fully incorporated into prices and interest rates, and as a consequence have no real effects. In addition, the fact that expectations concern the future actions of another agent raises issues of commitment and credibility.

Although not the focus of our analysis, it is useful for reference purposes to derive the Nash and Ramsey equilibria of the rational expectations version of the model. Knowing these equilibria will help us to interpret the learning dynamics.

Not coincidentally, our model turns out to be nearly isomorphic to the closed-economy model of Svensson (1997), whose focus was on inflation-targeting procedures. The main difference is that our central bank cares about the nominal exchange rate rather than the inflation rate. Also, because we are not concerned with the issue of "stabilization bias" per se, we assume that the central bank does not observe the current period Phillips Curve shock, so that unlike Svensson (1997), policy cannot be made contingent on the realization of v_{1t}. Finally, a minor difference is that our output equation contains an intercept. The presence of an intercept is unimportant with rational expectations, but acquires some significance with learning. In what follows we shall take full advantage of these similarities and refer the interested reader to Svensson (1997) for a more detailed discussion.

3.1. Nash Equilibrium

As is typically the case in dynamic settings, there are potentially many Nash equilibria of our model, depending on the exact nature of history dependence. We follow the standard practice in macroeconomics by focusing on Markov perfect equilibria. Denoting the parameter vector by $\theta = (\beta, \bar{y}, s^*, y^*, \lambda, \rho, \alpha)$, we obtain the following result:

PROPOSITION 3.1. *If Assumptions 3.2 and 3.3 (given below) are satisfied, there exists a unique stationary Markov perfect Nash equilibrium given by*

$$s_t = h_0(\theta) - h_1(\theta)(y_{t-1} - \bar{y}) \tag{10}$$

$$y_t = (1-\rho)\bar{y} + \rho y_{t-1} + \sigma_1 v_{1t}, \tag{11}$$

where

$$h_0(\theta) = s^* + \frac{\alpha\lambda(y^* - \bar{y})}{1 - \beta(\rho + \alpha h_1(\theta))} \tag{12}$$

$$h_1(\theta) = \frac{1 - \beta\rho^2 - \sqrt{(1-\beta\rho^2)^2 - 4\lambda\alpha^2\beta\rho^2}}{2\alpha\beta\rho}. \tag{13}$$

The proof is omitted because it involves minor adaptations of Svensson (1997). Note that if we are to guarantee a real-valued $h_1(\theta)$ function and a finite $h_0(\theta)$ function we must adopt the following two assumptions:

Assumption 3.2. The output weight in the government's objective function satisfies the inequality $\lambda < (1-\beta\rho^2)^2/4\alpha^2\beta\rho^2$.

Assumption 3.3. The output persistence parameter satisfies the (implicit) inequality $\rho < \beta^{-1} - \alpha h_1(\theta)$.

It can be verified that, as long as $0 < \rho < 1$, the feedback coefficient $h_1(\theta)$ has the same sign as the Phillips curve slope, α. Hence, given our assumption that adverse balance sheet effects dominate the relaxed borrowing constraint effect, so that $\alpha < 0$, our model predicts that the government will attempt to stabilize output by *appreciating* the currency during a recession. The contractionary effects of a higher real interest rate are more than offset by the increased net worth of firms caused by the decreased domestic currency value of foreign debt. Of course, with rational expectations, agents in the private sector are fully aware of this incentive and factor it into their expectations, so that at equilibrium the government is unsuccessful in its stabilization efforts.

3.2. Ramsey Equilibrium

The Nash equilibrium assumes that the government is unable to commit to a policy, and consequently must take the expectations of the private sector as given.

It is also of interest to consider a Ramsey equilibrium, where commitment is allowed. In a Ramsey equilibrium the government acts as a Stackelberg leader and takes advantage of its ability to shape private sector beliefs. However, unlike the case in Svensson's model, the government here has no ability to react to supply shocks after the private sector has formed its forecasts. As a result, it realizes that it can do no better than to set the exchange rate equal to its target, and the only sense in which it "shapes" private sector beliefs is by committing to this non-state-contingent policy. This gives us

PROPOSITION 3.4. *If the government can commit to an exchange rate policy, there exists a Ramsey equilibrium characterized by the following system:*

$$s_t = s^* \tag{14}$$

$$y_t = (1-\rho)\bar{y} + \rho y_{t-1} + \sigma_1 v_{1t}. \tag{15}$$

Hence, in the rational expectations version of our model, exchange rate fluctuations reflect a lack of commitment. Comparing the Nash outcome in (10) and (11) to the Ramsey outcome in (14) and (15) makes clear that a government lacking credibility suffers from an "appreciation bias." That is, the currency will be suboptimally strong. As usual, the extent of the bias depends on the degree to which the natural rate of output, $\bar{y}$, falls short of the target output rate, y^*, the weight on output fluctuations in the objective function, λ, and the importance of adverse balance sheet effects, α. Also, as stressed by Svensson, the discrepancy between the Nash and Ramsey exchange rates is increasing in the persistence of output. Greater output persistence enhances the temptation to engage in surprise appreciations, because it implies more "bang for the buck." This appreciation bias will play a leading role in our subsequent analysis of learning dynamics.

4. ADAPTIVE LEARNING

In principle, we could at this point proceed to estimate and/or simulate the model given in the previous section by (10), (11) or (14), (15). Doing this might reveal some interesting effects of commitment on exchange rate dynamics. For example, our model predicts that commitment produces a smoother exchange rate process. Also, it should be noted that even with rational expectations the discretionary equilibrium will generate *some* exchange rate dynamics, because the dynamics in y_t generate a time-varying incentive to depreciate, which then leads to expected and actual depreciations.[10] However, these dynamics will not look anything like currency crises. They will be smooth and episodic. If the underlying shocks are symmetrically distributed, the (linearized) rational expectations equilibria will have a hard time replicating the sharp nonlinearities that, almost by definition, characterize observed currency crises.

As noted in the Introduction, the conventional strategy for introducing nonlinearities is to exploit the nonlinearity of the budget constraint in models with borrowing constraints. This can produce multiple equilibria and open the door to sunspot fluctuations. Our paper pursues an alternative strategy. We inject nonlinearity by introducing adaptive feedback between beliefs and outcomes. This puts us in a different space. Rather than focus on switches between multiple steady states, our analysis focuses on the tail events of a single stochastic process.

We regard a learning approach to currency crises as more attractive and persuasive than a multiple equilibrium/sunspot approach, despite the fact that our learning algorithm is only "boundedly rational" whereas sunspot equilibria are "rational." Of course, the downside to a boundedly rational learning approach is that it requires us to specify *two* models, one for the true structure of the economy and one for the beliefs or perceptions of the agents in the economy. The great virtue of rational expectations is that we only need to specify one. This means that the persuasiveness of our model will partly depend on a priori assessments of the plausibility of agents' beliefs.

With this in mind, we try to be fairly unrestrictive and realistic in specifying the beliefs of the government and the private sector. In particular, we assume that the government's perceived model takes the following form:

$$y_t = \gamma_0 + \gamma_1 y_{t-1} + \gamma_2 s_t + u_t. \tag{16}$$

Comparing this to the actual model in (7), one can see that the government's model contains a subtle misspecification. It fails to properly account for the expectations of the private sector. As a result, the evolving beliefs of the private sector manifest themselves as parameter drift in the government's model. Following Sargent (1999) and CWS (2002), we will assume that the government responds to this drift by placing more emphasis on recent data when updating the parameters of its model.

Given the perceived model in (16), the government proceeds to minimize the loss function in (8). This is a standard LQR optimization problem, with the solution

$$s_t^p = g_0(\gamma) + g_1(\gamma) y_{t-1}, \tag{17}$$

where

$$g_0(\gamma) = \frac{s^* + (\lambda y^* - \beta p_1)\gamma_2 - \gamma_0\gamma_2(\lambda + \beta p_2)}{1 + \gamma_2^2(\lambda + \beta p_2)} \tag{18}$$

$$g_1(\gamma) = \frac{-\gamma_1\gamma_2(\lambda + \beta p_2)}{1 + \gamma_2^2(\lambda + \beta p_2)}, \tag{19}$$

where p_1 and p_2 are value function coefficients,

$$V(y_{t-1}) = p_0 + p_1 y_{t-1} + p_2 y_{t-1}^2,$$

and are given by

$$p_1 = \frac{g_1(\gamma)(g_0(\gamma) - s^*) + (\gamma_1 + g_1(\gamma)\gamma_2)[(\gamma_0 + g_0(\gamma)\gamma_2)(\lambda + \beta p_2) - \lambda y^*]}{1 - \beta(\gamma_1 + g_1(\gamma)\gamma_2)} \tag{20}$$

$$p_2 = -\gamma_1 g_1(\gamma)/\gamma_2. \tag{21}$$

As a check, note that the model studied by CWS (2002) corresponds to the parameter settings $s^* = y^* = \rho = 0$ and $\lambda = 1$. This implies that $p_1 = p_2 = \gamma_1 = 0$, which in turn implies that $g_1(\gamma) = 0$ and $g_0(\gamma) = -\gamma_0\gamma_2/(1 + \gamma_2^2)$. This is the same decision rule as in CWS.

Following CWS, we assume that the actual market exchange rate is equal to the government's planned exchange rate, s_t^{p}, plus an i.i.d. shock, which captures random implementation errors or high-frequency money demand shocks. Thus, we have

$$s_t = s_t^{\mathrm{p}}(\gamma) + \sigma_2 v_{2t}, \tag{22}$$

where $v_{2t} \sim N(0, 1)$.

Now, in our model the only "action" taken by the private sector in response to new information is to revise its expectations of next period's exchange rate.[11] Hence, the private sector just needs to formulate a model of the exchange rate. To be consistent with the government's beliefs and actions, we assume that it takes the following form:

$$s_t = \delta_0 + \delta_1 y_{t-1} + u_{2t}. \tag{23}$$

Under CWS's Fed watcher assumption, we would have $\delta_0 = g_0(\gamma)$ and $\delta_1 = g_1(\gamma)$.

If we now use these perceived laws of motion to evaluate the expectations in (7), it can readily be verified that the *actual* law of motion takes the form

$$y_t = (1 - \rho)\bar{y} - \alpha\delta_0 + (\rho - \alpha\delta_1)y_{t-1} + \alpha s_t + \sigma_1 v_{1t}. \tag{24}$$

At this point we confront the issue of the consistency between the government's perceived model in (16) and the actual model determined by those perceptions, given by (24). Unlike applications of adaptive expectations in the 1960s, we are going to demand that the government's beliefs be in some sense consistent with reality. This brings us to the notion of a self-confirming equilibrium.

4.1. Self-Confirming Equilibrium

The misspecification of the government's model prevents it from learning the rational expectations equilibrium derived in Section 3.[12] Despite this handicap, the government and the private sector both act purposefully to eliminate systematic forecast errors. They do this by choosing the parameters of their perceived models to best fit the data. Because their models include intercepts, agents will be successful at avoiding systematic forecast errors. So at least in this sense our model is not vulnerable to the kind of criticism that was leveled at the original applications of adaptive expectations in the 1960s. Still, the misspecification does mean that agents can miss the data's higher order moments and, as a consequence, there will in general be patterns in the forecast errors.

Although these patterns could be discovered if the agents were to explore alternative model specifications, we rule out this kind of experimentation. Following Sargent (1999), we adopt a weaker notion of equilibrium, which is well suited to models of boundedly rational learning. This equilibrium concept is called a "self-confirming equilibrium." A self-confirming equilibrium is weaker than a rational expectations equilibrium in the sense that it merely requires beliefs to be confirmed along the equilibrium path. Beliefs about off-equilibrium-path events can be arbitrary. As noted by Sargent (1976) and Hansen and Sargent (2001), off-equilibrium path play relates to the presence of regime changes. Given the historical data record, our agents' beliefs can always be made consistent with the data. However, beliefs will not be optimal in the rational expectations sense of fully conforming to the actual data-generating process. In particular, they are vulnerable to out-of-sample regime changes.[13]

If this restricted notion of optimality is defined in terms of minimizing the variance of one-step-ahead forecast errors, then the following least-squares normal equations characterize a self-confirming equilibrium,

$$E\left\{\begin{pmatrix} y_{t-1} \\ 1 \end{pmatrix}[s_t - \delta_0 - \delta_1 y_{t-1}]\right\} = 0 \tag{25}$$

$$E\left\{\begin{pmatrix} y_{t-1} \\ s_t \\ 1 \end{pmatrix}[y_t - \gamma_0 - \gamma_1 y_{t-1} - \gamma_2 s_t]\right\} = 0, \tag{26}$$

where the expectations are evaluated using the distribution implied by the true data-generating process in (24). Parameter values that satisfy these equations have the property that agents do not have an incentive to revise the parameters of their models.

We can now state a more precise definition of our equilibrium concept.

DEFINITION 4.1. *A self-confirming equilibrium is a collection of regression coefficients* $(\hat{\delta}_0, \hat{\delta}_1, \hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2)$ *and an exchange rate policy* $s_t^p = g_0(\gamma) + g_1(\gamma) y_{t-1}$ *such that when* y_t *is governed by* (24)*, the regression coefficients satisfy the least-squares orthogonality conditions in* (25) *and* (26).

A self-confirming equilibrium is characterized in the following proposition.

PROPOSITION 4.2. *Given the perceived laws of motion in equations* (16) *and* (23) *and Assumption 3.2, a unique self-confirming equilibrium exists if* $|\rho| < 1$, *which is characterized by the recursive system of equations*

$$\delta_0 = g_0(\gamma) \tag{27}$$

$$\delta_1 = g_1(\gamma) \tag{28}$$

$$\gamma_2 = \alpha \tag{29}$$

$$0 = \beta\rho\gamma_1^2 - (1 + \beta\rho^2)\gamma_1 + \rho(1 + \lambda\alpha^2) \tag{30}$$

$$\widetilde{g}_1 = \frac{-\left[\beta\gamma_1^2 - \left(1 + \lambda\gamma_2^2\right)\right] - \sqrt{\left[\beta\gamma_1^2 - \left(1 + \lambda\gamma_2^2\right)\right]^2 + 4\beta\lambda\gamma_1^2\gamma_2^2}}{2\beta\gamma_1\gamma_2} \tag{31}$$

$$\gamma_0 = \frac{\bar{y}(1-\rho)\left[1-\beta\gamma_1-\beta\widetilde{g}_1\gamma_2(1+\beta)+\gamma_2^2(\lambda+\beta p_2)\right]-\alpha[s^*(1-\beta\gamma_1)+y^*\lambda\alpha]}{1-\beta\gamma_1-\beta(1+\beta)\widetilde{g}_1\gamma_2}, \tag{32}$$

where p_2 *is given by* (21), *and* $g_0(\gamma)$ *and* $g_1(\gamma)$ *are given by* (18) *and* (19).

Proof. See the Appendix.

The fact that the self-confirming equilibrium is unique here differentiates our analysis from that of Kasa (2004), where crises represent switches between multiple self-confirming equilibria. This paper shows that crises can occur even with a unique self-confirming equilibrium.

The above definition and characterization of a self-confirming equilibrium are stated in terms of population moments. In practice, agents do not know these moments. They must be estimated from the data. Following Sargent (1999) and CWS (2002), we assume that agents do this via a recursive least-squares procedure. Letting $\gamma_t' = (\gamma_{0t}, \gamma_{1t}, \gamma_{2t})$ and $z_{g,t} = (1, y_t, s_{t+1})'$, we can write the government's learning algorithm as

$$\gamma_t = \gamma_{t-1} + a_g R_{gt-1}^{-1} z_{g,t-1}(y_t - z_{g,t-1}'\gamma_{t-1}) \tag{33}$$

$$R_{gt} = R_{g,t-1} + a_g(z_{g,t-1}z_{g,t-1}' - R_{g,t-1}). \tag{34}$$

The private sector uses an analogous learning algorithm,

$$\delta_t = \delta_{t-1} + a_p R_{pt-1}^{-1} z_{p,t-1}(s_t - z_{p,t-1}'\delta_{t-1}) \tag{35}$$

$$R_{pt} = R_{p,t-1} + a_p(z_{p,t-1}z_{p,t-1}' - R_{p,t-1}), \tag{36}$$

where $\delta_t' = (\delta_{0t}, \delta_{1t})$ and $z_{p,t} = (1, y_t)'$.

There are several points to notice about these algorithms. First, although they are apparently ad hoc, one can interpret these algorithms as an approximation to a conventional (i.e., Bayesian) Kalman filter under a particular prior specification [see, e.g., Sargent and Williams (2005)]. Second, it is important to keep in mind that the data processes that appear on the right-hand sides are the *true* data-generating processes in (22) and (24). This makes the model self-referential and complicates the analysis of the learning dynamics. Not only do outcomes affect beliefs via the learning algorithms in (33)–(36), but also beliefs feed back to influence the observed data via equations (22) and (24). Third, the learning algorithms imply that beliefs are changing each period. However, when solving their optimization and forecasting problems, agents act as if their beliefs do not change. This is of course "irrational," but one could argue that it is at least as descriptively accurate as assuming that agents base current actions on forecasts of their future beliefs. Finally, the crucial parameters in these algorithms are the two gain parameters, a_g and a_p. These dictate how responsive beliefs are to new information. In a simple least-squares procedure, the gains would decrease to zero at rate t^{-1}, reflecting the fact that each innovation adds less and less information relative to the accumulated stock of prior experience. However, as noted in the Introduction, we assume that agents pay more attention to recent data. They do this because they suspect that the environment is nonstationary. This is accomplished by constraining the gain parameters to be (small) constants. This effectively discounts old data at rate $(1 - a_i)$ and allows agents to remain alert to potential structural breaks. Interestingly, it turns out that constant-gain learning algorithms can produce in an endogenous and self-confirming manner exactly the kind of instability that they are designed to guard against.

4.2. Stability

We can now analyze the dynamic system consisting of the belief revision processes (33)–(36) and the data-generating processes in (22) and (24). This is not easy. Besides being a nonlinear, dynamic, stochastic system, with all the attendant difficulties these features entail, the system is also self-referential. Not only do beliefs respond to the data, but also the data respond to beliefs. As a result, we must resort to nonstandard methods of analysis.

The key idea behind these methods is to take advantage of the fact that for small gain parameters, (a_g, a_p), beliefs and the data operate on different time scales. Beliefs are a slow process, and the data are a fast process. This opens the door to so-called *singular perturbation* methods. For us, this means we can study the evolution of beliefs by first averaging over the data *for fixed beliefs*. When the gain decreases at rate t^{-1}, as with recursive least squares, this time-scale separation strategy produces a single deterministic ODE that fully characterizes the limiting behavior of beliefs. However, with constant gain, beliefs do not converge to a fixed point. Instead, they converge in a distributional sense. In particular, they converge (weakly) to a diffusion process. The asymptotic (with respect to the gain) behavior of this diffusion process can be characterized with *two* ODEs. One is the conventional mean dynamics, describing the efforts of agents to eliminate systematic forecast errors. This pulls the system toward the SCE. More interestingly, the second ODE describes the path of rare, but recurrent *escapes* from the SCE. This second ODE is the solution of a *deterministic* optimal control problem, which has the interpretation of minimizing the cost, in probabilistic terms, of escaping from the SCE.

To see how this works, let us begin by stacking up (33)–(36). This can be written

$$\begin{bmatrix} \gamma_{t+1} \\ \delta_{t+1} \end{bmatrix} = \begin{bmatrix} \gamma_t \\ \delta_t \end{bmatrix} + a_{\mathrm{g}} \left[f(\gamma_t, \delta_t, \lambda) + \mu_{t+1} \right], \tag{37}$$

where μ_{t+1} is a martingale difference, and $\lambda = a_{\mathrm{p}}/a_{\mathrm{g}}$, which measures the private sector's speed of learning relative to the government's. The associated mean dynamics ODE is

$$\begin{bmatrix} \dot{\gamma} \\ \dot{\delta} \end{bmatrix} = f(\gamma, \delta, \lambda). \tag{38}$$

Note this does not depend on the data. We have averaged out over the data with respect to their stationary distribution given current beliefs. Time-scale separation allows us to do this.

It can readily be verified that an SCE is just a stationary solution of (38). As always, one would like to know whether this equilibrium is stable in some sense. The standard definition of stability in the recursive learning literature is "E-stability." An equilibrium is E-stable if and only if the stationary solution of the mean dynamics are locally stable.

DEFINITION 4.3. *A stationary solution* (δ^*, γ^*) *of* (38) *is locally stable if and only if the eigenvalues of the Jacobian of* $f(\gamma, \delta, \lambda)$ *have negative real parts at* (δ^*, γ^*).

The central result of recursive learning models is that if a stationary solution of (38) is locally stable, then the recursive learning algorithm (37) locally converges to it.[14] Because there is a unique self-confirming equilibrium, (38) must have a unique stationary solution. To see whether (37) converges to the self-confirming equilibrium, it suffices to check whether the self-confirming equilibrium is locally stable, and hence E-stable.

PROPOSITION 4.4. *If* $0 < \rho < 1$, *the unique self-confirming equilibrium is (locally) stable.*

Proof. See the Appendix.

The proof makes clear that $0 < \rho < 1$ is only a sufficient condition for stability, not a necessary condition. This is not a serious limitation, however, because negative values of ρ are not empirically relevant.

4.3. Escape Dynamics

Equation (38) only characterizes the *mean* of the distribution of the sample paths. With a (small) constant gain, parameter estimates experience rare but recurrent escapes from the SCE. Large deviation methods provide information about these tail events. For example, they tell us where the estimates escape to, and how often they do so. These escape routes and escape times are the outcome of a (highly nonlinear) control problem. The key input to this control problem is the Legendre transform of the log-moment generating function of the above martingale difference noise sequence, μ_{t+1}. In general, this can be very difficult to calculate. However, in an LQG model such as ours, the computations are drastically simplified. Williams (2001) provides computational algorithms.

4.4. Two-Sided Learning Dynamics

While our model has much in common with those of Sargent (1999) and CWS (2002), there is one potentially important difference. To generate more realistic output dynamics, we abandon their "Fed watcher" assumption and assume that the private sector must also learn. In this section, we analyze the consequences of this extension. Perhaps surprisingly, it has no effect on the large deviations properties of our model. To see why, we need explicit formulas for f and μ_t:

$$f(\gamma_t,\delta_t,\lambda)=\begin{bmatrix} R_{gt}^{-1}\begin{bmatrix} A_{0,t}+A_{1,t}\bar{y}(\gamma_t,\delta_t) \\ A_{0,t}\bar{y}(\gamma_t,\delta_t)+A_{1,t}\mathrm{E}(y^2) \\ (g_0(\gamma_t)+g_1(\gamma_t)\bar{y}(\gamma_t,\delta_t))(A_{0,t}+A_{1,t}\bar{y}(\gamma_t,\delta_t)) \\ +(\alpha-\gamma_2)\sigma_2^2+A_{1,t}g_1(\gamma_t)\sigma_y^2 \end{bmatrix} \\ R_{pt}^{-1}\begin{bmatrix} (g_0(\gamma_t)-\delta_0)+(g_1(\gamma_t)-\delta_1)\bar{y}(\gamma_t,\delta_t) \\ (g_0(\gamma_t)-\delta_0)\bar{y}(\cdot)+(g_1(\gamma_t)-\delta_1)\mathrm{E}(y^2) \end{bmatrix} \end{bmatrix} \tag{39}$$

and

$$\mu_{t+1}=\begin{bmatrix} R_{gt}^{-1}\begin{bmatrix} (\alpha-\gamma_2)\sigma_2 v_{2,t+1}+\sigma_1 v_{1,t+1}+A_{1,t}(y_t-\bar{y}(\gamma_t,\delta_t)) \\ A_{0,t}(y_t-\bar{y}(\gamma_t,\delta_t))+A_{1,t}(y_t^2-\mathrm{E}(y^2)) \\ +\sigma_1 y_t v_{1,t+1}+(\alpha-\gamma_2)\sigma_2 y_t v_{2,t+1} \\ (\alpha-\gamma_2)\sigma_2^2(v_{2,t+1}^2-1)+A_{1,t}g_1(\gamma_t)(y_t^2-\mathrm{E}(y^2)) \\ +\sigma_1\sigma_2 v_{1,t+1}v_{2,t+1}+\Gamma_{t+1} \end{bmatrix} \\ R_{pt}^{-1}\begin{bmatrix} (g_1(\gamma_t)-\delta_1)(y_t-\bar{y}(\cdot))+\sigma_2 v_{2,t+1} \\ (g_0(\gamma_t)-\delta_0)(y_t-\bar{y}(\cdot))+(g_1(\gamma_t)-\delta_1)(y_t^2-\mathrm{E}(y^2)) \\ +\sigma_2 v_{2,t+1} \end{bmatrix} \end{bmatrix}, \tag{40}$$

where $A_{0,t}$, $A_{1,t}$, and Γ_{t+1} are defined as

$$\begin{aligned} A_{0,t}&=(1-\rho)\bar{y}-\alpha\delta_{0t}-\gamma_{0t}+(\alpha-\gamma_{2t})g_0(\gamma_t) \\ A_{1,t}&=(\rho-\alpha\delta_{1t}-\gamma_{1t})+(\alpha-\gamma_{2t})g_1(\gamma_t) \\ \Gamma_{t+1}&=(g_0A_1+g_1A_0)(y_t-\bar{y}(\cdot))+\sigma_1(g_0+g_1y_t)v_{1,t+1} \\ &\quad+\sigma_2[(\alpha-\gamma_2)(g_0+g_1y_t)+A_0+A_1y_t]v_{2,t+1} \end{aligned}$$

and $\bar{y}(\gamma_t,\delta_t)$ and σ_y^2 are the stationary mean and variance of y_t *given* γ_t and δ_t.

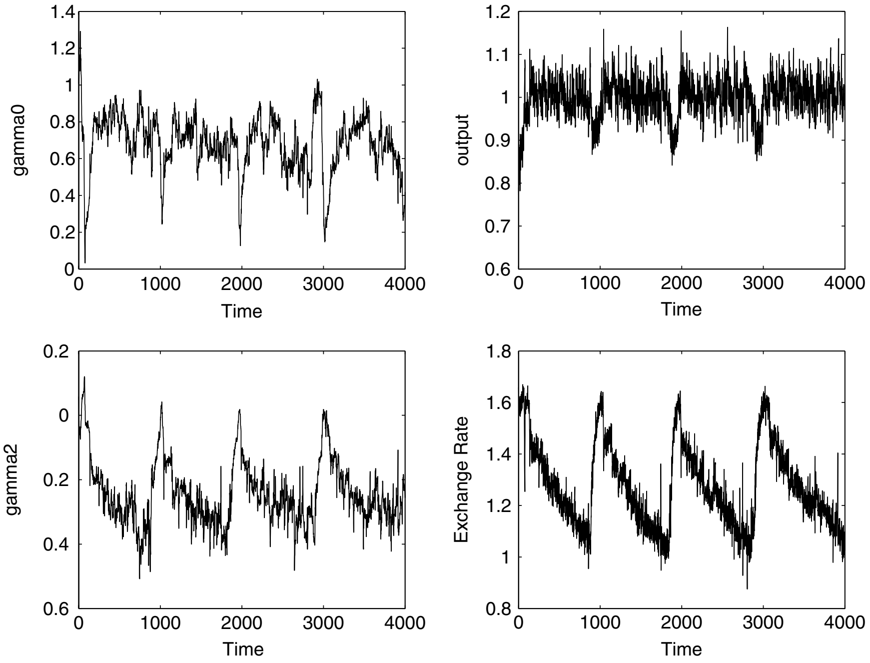

FIGURE 1. Representative sample paths when $a_g = a_p = 0.04$. The self-confirming equilibrium exchange rate is 1.0 and the Ramsey rate is 1.6.

The noise term in (40) reveals two interesting things. First, notice that it is *not* Gaussian, even when the underlying shocks are. Second, notice that it is a stochastically singular process. In particular, notice that the final two rows, giving the innovations in private sector beliefs, are completely determined by the first three rows. As a result, the beliefs of the private sector do not contribute to the escape dynamics. Intuitively, this derives from the fact that the private sector takes no "actions"; here, it merely tries to forecast the actions of the government. Appendix C contains a more detailed discussion of this point.

4.5. Simulations

Figure 1 reports representative sample paths from the model. The key parameter values are as follows: (1) $a_p = a_g = 0.04$, which implies a half-life of data relevance of about 17 time periods. (2) $\alpha = -0.3$, which reflects our assumption that adverse balance sheet effects dominate liquidity and competitiveness effects. Given our assumed value of $\rho = 0.7$, this is roughly consistent with a 30% foreign debt/GDP ratio.[15] (3) $\lambda = 1.5$, which implies that output fluctuations are more costly than exchange rate fluctuations. (4) $\bar{y} = 1.0$ and $y^* = 1.2$, which reflect the usual assumption in these models that the target output level exceeds the natural rate (perhaps due to imperfect competition or labor market distortions). (5) $\sigma_1^2 = 0.0003$ and $\sigma_2^2 = 0.0001$, which imply that real shocks are more volatile than nominal shocks. (6) $s^* = 1.6$, which is arbitrary.

At least in a qualitative sense, Figure 1 shows that our model generates output and exchange rate paths that resemble those in many crisis-prone countries. First, notice that the exchange rate path is highly *asymmetric*. There are prolonged periods of gradual appreciation, followed by sharp depreciations. The appreciation phase reflects the government's incentive to keep the exchange rate strong in the presence of adverse balance sheet effects. Second, crises are *recurrent*. In this particular case, they occur approximately once every 1,000 periods, that is, about once every 4–5 years if the time unit is a day, or about once every 20 years if the time unit is a week. Increasing the gain parameters or the shock variances increases the frequency of crises. Third, *crises cause recessions*, with output typically falling by about 10% during a crisis. These recessions will be more or less persistent, depending on the value of ρ.

Notice that crises are intimately connected to the evolving beliefs of the government. The two plots on the left-hand side of Figure 1 depict the paths of the coefficient estimates. The cyclical pattern of the coefficient estimates mirrors the cyclical pattern of the exchange rate. What is happening here is that the government is vacillating between Sargent's (1976) two observationally equivalent ways of interpreting the data. Crises occur when the government confuses the natural rate properties of the model with the apparent absence of balance sheet effects. As in CWS, this happens when the shock to the Phillips curve, v_{1t}, and the shock to the exchange rate, v_{2t}, move together. Ceteris paribus, positive realizations of v_{2t} produce equal contemporaneous movements in the exchange rate. From the expectational Phillips curve, this produces negative shocks to output. However, if at the same time there are positive v_{1t} shocks, these will offset the effect of the exchange rate shock, making it appear as if output does not respond to the exchange rate. This leads the government to reduce its estimate of balance sheet effects, which makes it more willing to tolerate a depreciation.

5. RELATION TO THE EXISTING LITERATURE

The central actor in our account of currency crises is a government policy maker contending with model uncertainty. Rather than suppose that he is an omniscient Bayesian, capable of reducing his uncertainty about the world to a finite-dimensional parameterization, we suppose instead that he is a workaday econometrician, who adopts tentative and provisional models and is unaware of his own influence on the data-generating processes about which he is learning. Despite his shortcomings, we claim he is still quite smart. He revises his model to improve its fit and eliminate systematic forecast errors, and he solves a dynamic programming problem to compute a policy that optimally trades-off exchange rate and output fluctuations. His crucial mistake is not to pay adequate attention to the identification problem. He could improve the specification of his model by following the advice of Sargent (1976) and thinking "off the equilibrium path."

Although we believe this story captures an important element in many actual crisis episodes, it is admittedly quite unconventional. The conventional wisdom is that governments are forced off pegs after they deplete their reserves. In contrast, our government chooses to devalue! In our view, this is a semantic distinction. As Obstfeld and Rogoff (1995) note, governments always have the option to maintain a peg through high enough interest rates, explicit default, etc. The real question is how high a cost the government is willing to pay in order to sustain it. We agree with Obstfeld and Rogoff that the right way to think about collapsing pegs is by studying the *choices* that governments make.

Still, it is undoubtedly the case that by treating foreign debt as exogenous, and thereby abstracting from the actions of the private sector, we are throwing out an important piece of the puzzle. As Chang and Velasco (2006) note, the *interaction* between foreign debt decisions and optimal exchange rate policy has the structure of a coordination game and can easily generate multiple equilibria. Although this would certainly enhance the realism of our model, it would not change the basic tenor of its results. For example, Kasa (2004) shows that escape dynamics can generate endogenous switches between multiple SCE.

Perhaps the most closely related prior work is by Marcet and Nicolini (2003). Like us, they show how adaptive learning dynamics can generate recurrent crises (hyperinflations in their case). The key mechanism in their analysis is a state-contingent gain sequence. During tranquil periods, agents use a least-squares/decreasing-gain sequence. However, during turbulent periods, when forecast errors exceed an exogenous threshold, agents switch to a constant-gain algorithm. However, we believe our analysis possesses some important advantages. First, Marcet and Nicolini's model is not amenable to large deviations methods. Rather than exploiting nonlinear feedback to endogenously propel the system away from the SCE, their model requires large shocks to kick the system into an unstable region of the parameter space. Second, large deviations methods enable us, at least in principle, to characterize *analytically* the statistical properties of crises and thereby relate them to the underlying parameters of the economy. In contrast, Marcet and Nicolini (2003) must resort to simulations. Third, our model can explain why crises are contractionary. Marcet and Nicolini's model cannot address this issue.

6. CONCLUSION

This paper extends the third-generation currency crisis literature by modeling the high-frequency dynamics of currency crises. We do this by explicitly modeling the evolution of beliefs. We share the opinion of other contributors to this literature that beliefs lie at the heart of currency crises. However, rather than regarding these beliefs as responding in an implausibly coordinated way to exogenous sunspots, we regard beliefs as responding adaptively to recent experience. Despite the adaptive nature of beliefs, we show that the nonlinearity induced by self-referential feedback between beliefs and outcomes can produce what look like switches between multiple equilibria. However, in our model, currency crises are *not* switches between multiple equilibria. They reflect the escape dynamics of a unique equilibrium stochastic process.

Our model attributes currency crises to government miscalculation and model misspecification. Crises occur when the government underestimates the contractionary effects of currency depreciation. Unanticipated depreciations are contractionary due to the presence of unhedged foreign currency debt and its adverse balance sheet effects. We assume that the government is unsure about the economy's exposure to these balance sheet effects and must revise its beliefs about them recursively as it witnesses the economy's response to its exchange rate policy. When doing this, the government uses a misspecified model, which misinterprets the role of private sector expectations.

The importance of uncertainty about balance sheet effects is now widely appreciated among policy makers. Garber (1999) argues persuasively that the increasing importance of derivatives makes it easier for market participants to circumvent regulations designed to limit foreign exchange exposure, while at the same time making it more difficult for governments to detect this activity. Draghi et al. (2003) argue that often governments unwittingly expose themselves to foreign currency risks due to the state-contingent nature of many government policies. However, there has been little formal analysis of this issue in the currency crisis literature, perhaps due to its heavy reliance on the rational expectations hypothesis. Our paper is the first to show how learning about balance sheet effects can generate currency crises.

There are several possible extensions of our analysis. One commonly observed feature of currency crises is their tendency to be *contagious*. Crises often spread across countries. A recent paper by Ellison et al. (2006) develops a two-country version of our model and shows that escape dynamics can be contagious, even between countries that are weakly linked by fundamentals. Another extension would be to introduce an explicit preference for robustness [see Hansen and Sargent (2001)]. The government in our model is alert to model misspecification, but when the time comes to formulate an exchange rate policy, it ignores model uncertainty. It would be interesting to see how a preference for robustness in control would influence the dynamics of currency crises. Tetlow and von zur Muehlen (2004) study this issue in the context of Sargent's (1999) model. Finally, one could argue that constant-gain learning is based on an assumption (slow parameter drift) that is violated by the model's escape dynamics. In Cho and Kasa (2005a), we show that a standard testing and model validation process can provide a more convincing behavioral foundation for constant gain learning. In Cho and Kasa (2005b) we show that validation dynamics can produce currency crises, but a complete analysis remains the subject of future research.

NOTES

1. See Flood and Marion (1999) for a survey. Krugman (1996) and Morris and Shin (1998) express skepticism about the relevance of multiple equilibria.

2. Leading papers include Aghion et al. (2001), Cespedes et al. (2004), Krugman (1999), and Schneider and Tornell (2004).

3. Burnside et al. (2001) and Aghion et al. (2001) provide a variety of exposure estimates for both Asia and Mexico.

4. Even those who advocate a fundamentals-based/first-generation account of recent currency crises often resort to sunspots when it comes to explaining their timing. See, e.g., Burnside et al. (2004).

5. Of course, learning and multiple equilibria are not mutually exclusive. Woodford (1990) shows that sunspot equilibria can be learned. Kasa (2004) introduces adaptive learning into the escape clause model of Obstfeld (1997) and shows that learning dynamics, rather than sunspots, can generate switches between multiple steady states.

6. See Bray and Kreps (1987) for a discussion of some conceptual problems associated with Bayesian learning. Kreps (1998) provides arguments in favor of a boundedly rational approach to learning, which is quite similar to ours. He calls it "anticipated utility." Cogley and Sargent (2005) compare active Bayesian learning with passive anticipated utility learning. In their model the two approaches produce very similar outcomes. However, they caution that there can be significant differences when risk aversion is important.

7. This specification error is subtle in the sense spelled out by Sargent (1976); viz. given any historical data record, the two models are *observationally equivalent*.

8. See, e.g., Dooley (2000), Burnside et al. (2001), and Schneider and Tornell (2004).

9. Note that ABB are able to ignore the risk premium because they only consider unanticipated one-time shocks. In general, the risk premium will be a nonlinear function of the net worth/capital ratio. See, e.g., Cespedes et al. (2004) for a detailed derivation of the risk premium in a model that is quite similar to ours.

10. In a similar model, Ireland (1999) argues that variation in "time-consistency bias" can explain the rise and fall of U.S. inflation during the 1970s and 1980s.

11. Note that this is likely a nonrobust feature of our model. For example, if foreign currency debt were *endogenous*, then the evolution of the private sector's beliefs would produce changes in the system via changes in the stock of foreign currency liabilities.

12. See Evans and Honkapohja (2001) for an extensive discussion of the circumstances under which it is and is not possible for agents to learn a model's rational expectations equilibria. As noted by a referee, it is not really the model's functional form that is the problem, but rather misinterpretation of observed parameter drift.

13. Evans and Honkapohja (2001) call this concept a restricted perceptions equilibrium.

14. For a formal analysis, see Marcet and Sargent (1989), Woodford (1990), and Evans and Honkapohja (2001).

15. From equation (2.6), $\alpha \approx -\rho(1+\bar{\mu})(1+i^*)(\bar{d}^*/\bar{y})$.

16. Note that we must select the smaller root to ensure $\lim_{\rho\to 0} g_1 = 0$.

REFERENCES

Aghion, P., P. Bacchetta, and A. Banerjee (2001) Currency crises and monetary policy in an economy with credit constraints. *European Economic Review* 45, 1121–1150.

Aghion, P., P. Bacchetta, and A. Banerjee (2004) A corporate balance-sheet approach to currency Crises. *Journal of Economic Theory* 119, 6–30.

Bogomolova, A., D. Kolyuzhnov, and S. Slobodyan (2005) Escape Dynamics: A Continuous-Time Approximation. Manuscript, CERGE-EI.

Bray, M. M. and D. Kreps (1987) Rational learning and rational expectations. In G. R. Feiwel (ed.), *Arrow and the Ascent of Modern Economic Theory*. New York: New York University Press.

Burnside, C., M. Eichenbaum, and S. Rebelo (2001) Hedging and financial fragility in fixed exchange rate regimes. *European Economic Review* 45, 1151–1194.

Burnside, C., M. Eichenbaum, and S. Rebelo (2004) Government guarantees and self-fulfilling speculative attacks. *Journal of Economic Theory* 119, 31–63.

Cespedes, L. F., R. Chang, and A. Velasco (2004) Balance sheets and exchange rate policy. *American Economic Review* 94, 1183–1193.

Chang, R. and A. Velasco (2006) Currency mismatches and monetary policy: A tale of two equilibria. *Journal of International Economics* 69, 150–175.

Chari, V. V. and P. J. Kehoe (2004) Financial crises as herds: overturning the critiques. *Journal of Economic Theory* 119, 128–150.

Cho, I. K. and K. Kasa (2005a) Learning and Model Validation. Manuscript, University of Illinois.

Cho, I. K. and K. Kasa (2005b) Model Validation Dynamics and Endogenous Currency Crises. Manuscript, University of Illinois.

Cho, I. K., N. Williams, and T. J. Sargent (2002) Escaping Nash inflation. *Review of Economic Studies* 69, 1–40.

Clarida, R., J. Gali, and M. Gertler (2001) Optimal monetary policy in open versus closed economies: An integrated approach. *American Economic Review* 91, 248–252.

Cogley, T. and T. J. Sargent (2005) Anticipated Utility and Rational Expectations as Approximations of Bayesian Decision Making. Manuscript, UC Davis.

Dooley, M. (2000) A model of crises in emerging markets. *Economic Journal* 110, 256–272.

Draghi, M., F. Giavazzi, and R. C. Merton (2003) Transparency, Risk Management and International Financial Fragility. NBER Working Paper No. 9806.

Ellison, M., L. Graham, and J. Vilmanen (2006) Strong contagion with weak spillovers. *Review of Economic Dynamics* 9, 263–283.

Evans, G. W. and S. Honkapohja (2001) *Learning and Expectations in Macroeconomics*. Princeton, NJ: Princeton University Press.

Flood, R. P. and N. Marion (1999) Perspectives on the recent currency crisis literature. *International Journal of Finance and Economics* 4, 1–26.

Garber, P. M. (1999) Derivatives in international capital flows. In M. Feldstein (ed.), *International Capital Flows*. Chicago: University of Chicago Press.

Hansen, L. P. and T. J. Sargent (2001) Acknowledging misspecification in macroeconomic theory. *Review of Economic Dynamics* 4, 519–535.

Ireland, P. N. (1999) Does the time-consistency problem explain the behavior of inflation in the United States? *Journal of Monetary Economics* 44, 279–291.

Jeanne, O. (2000) Foreign currency debt and the global financial architecture. *European Economics Review* 44, 719–727.

Kasa, K. (2004) Learning, large deviations and recurrent currency crises. *International Economic Review* 45, 141–173.

Kehoe, T. J. and D. K. Levine (1993) Debt-constrained asset markets. *Review of Economic Studies* 60, 865–888.

Kreps, D. M. (1998) Anticipated utility and dynamic choice. In *Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures, 1983–1997*. Cambridge: Cambridge University Press.

Krugman, P. (1996) Are currency crises self-fulfilling? In *NBER Macro Annual* Cambridge, MA: MIT Press.

Krugman, P. (1999) Balance sheets, the transfer problem, and financial crises. In P. Isard, A. Razin and A. Rose (eds.), *International Finance and Financial Crises: Essays in Honor of Robert P. Flood.* Dordrecht, the Netherlands: Kluwer.

Marcet, A. and J. P. Nicolini (2003) Recurrent hyperinflations and learning. *American Economic Review* 93, 1476–1498.

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

Morris, S. and H. S. Shin (1998) Unique equilibrium in a model of self-fulfilling currency attacks. *American Economic Review* 88, 587–597.

Obstfeld, M. (1994) The logic of currency crises. *Cahiers Economiques et Monetaires* 43, 189–213.

Obstfeld, M. (1997) Destabilizing effects of exchange rate escape clauses. *Journal of International Economics* 43, 61–77.

Obstfeld, M. and K. Rogoff (1995) The mirage of fixed exchange rates. *Journal of Economic Perspectives* 9, 73–96.

Sargent, T. J. (1976) The observational equivalence of natural and unnatural rate theories of macroeconomics. *Journal of Political Economy* 84, 631–640.

Sargent, T. J. (1993) *Bounded Rationality in Macroeconomics*. Oxford: Oxford University Press.

Sargent, T. J. (1999) *The Conquest of American Inflation*. Princeton, NJ: Princeton University Press.

Sargent, T. J. and N. Williams (2005) Impacts of priors on convergence and escapes from Nash inflation. *Review of Economic Dynamics* 8, 360–391.

Schneider, M. and A. Tornell (2004) Balance sheet effects, bailout guarantees and financial crises. *Review of Economic Studies* 71, 883–913.

Svensson, L. E. O. (1997) Optimal inflation targets, "Conservative" central banks, and linear inflation contracts. *American Economic Review* 87, 98–114.

Tetlow, R. and P. von zur Muehlen (2004) Avoiding Nash inflation: Bayesian and robust responses to model uncertainty. *Review of Economic Dynamics* 7, 869–899.

Tirole, J. (2003) Foreign borrowing: A dual- and common-agency perspective. *American Economic Review* 93, 1678–1702.

Williams, N. (2001) *Escape Dynamics in Learning Models*. Ph.D. Thesis, University of Chicago.

Woodford, M. (1990) Learning to believe in sunspots. *Econometrica* 58, 277–307.

APPENDIX A

PROOF OF PROPOSITION 4.2

First, by substituting (17) and (22) into (25), it is clear that a self-confirming equilibrium features $\delta_0 = g_0(\gamma)$ and $\delta_1 = g_1(\gamma)$. Hence, we can just focus on the government's beliefs. Next, by substituting (24) into (26) and defining the vectors $z_t = (1, y_{t-1}, s_t)'$ and $\gamma = (\gamma_0, \gamma_1, \gamma_2)'$, we can write the government's normal equations as

$$M_z(\gamma)[T(\gamma) - \gamma] = 0,$$

where $M_z(\gamma)$ is the 3×3 product–moment matrix, $E(zz')$, and $T(\gamma)$ defines a 3×1 vector of actual regression coefficients as a function of the perceived regression coefficients,

$$T(\gamma) = \begin{pmatrix} \bar{y}(1-\rho) - \alpha g_0(\gamma) \\ \rho - \alpha g_1(\gamma) \\ \alpha \end{pmatrix}, \tag{A.1}$$

where we have used (27) and (28) to substitute out private sector beliefs. From this we immediately conclude that $\gamma_2 = \alpha$. Hence, we have reduced the problem to two equations in γ_0 and γ_1. It turns out this system is recursive, because $g_1(\gamma)$ is independent of γ_0. So let us first solve for γ_1. The Bellman equation implies the following fixed point condition for p_2:

$$p_2 = g_1(\gamma)^2 + (\gamma_1 + g_1(\gamma)\gamma_2)^2(\lambda + \beta p_2). \tag{A.2}$$

From the government's policy function we have

$$\lambda + \beta p_2 = \frac{-g_1(\gamma)}{\gamma_2(g_1(\gamma)\gamma_2 + \gamma_1)}.$$

Substituting this into (A.2) yields the following quadratic equation characterizing the feedback coefficient $g_1(\gamma)$,

$$g_1(\gamma)^2 + \left(\frac{\gamma_1}{\gamma_2} - \frac{1 + \lambda\gamma_2^2}{\beta\gamma_1\gamma_2}\right) g_1(\gamma) - \lambda\beta^{-1} = 0, \tag{A.3}$$

with the solution[16]

$$g_1(\gamma) = \frac{-\left[\beta\gamma_1^2 - (1 + \lambda\gamma_2^2)\right] - \sqrt{\left[\beta\gamma_1^2 - \left(1 + \lambda\gamma_2^2\right)\right]^2 + 4\lambda\beta\gamma_1^2\gamma_2^2}}{2\beta\gamma_1\gamma_2}. \tag{A.4}$$

Next, substitute this into the second of the self-confirming equilibrium conditions, $T(\gamma) = \gamma$, which gives

$$\gamma_1 = \rho - \alpha g_1(\gamma).$$

Imposing the equilibrium condition $\gamma_2 = \alpha$ and simplifying then produces the quadratic equation given in (30). Assumption 3.1 is necessary for this equation to have real roots. In addition, to ensure that $\lim_{\rho\to 0}\gamma_1 = 0$ and $\lim_{\beta\to 0}\gamma_1 = \rho(1 + \lambda\alpha^2)$, we must select the smaller root. This implies uniqueness and yields

$$\gamma_1 = \frac{1 + \beta\rho^2 - \sqrt{(1 + \beta\rho^2)^2 - 4\beta\rho^2(1 + \lambda\alpha^2)}}{2\beta\rho}. \tag{A.5}$$

Finally, substituting the value function coefficients in (20) and (21) into the expression for $g_0(\gamma)$ in (18) and using the first self-confirming equilibrium condition, $\gamma_0 = \bar{y}(1 - \rho) - \alpha g_0(\gamma)$, produces a linear equation for γ_0, with unique solution given in (32).

APPENDIX B

PROOF OF PROPOSITION 4.4

From Evans and Honkapohja (2001), we know that stability requires the eigenvalues of the Jacobian of $T(\gamma)$, evaluated at the self-confirming equilibrium, to have real parts less than one. Using (A.1), the Jacobian takes the form

$$T_\gamma = \begin{pmatrix} -\alpha\dfrac{\partial g_0}{\partial \gamma_0} & -\alpha\dfrac{\partial g_0}{\partial \gamma_1} & -\alpha\dfrac{\partial g_0}{\partial \gamma_2} \\ 0 & -\alpha\dfrac{\partial g_1}{\partial \gamma_1} & -\alpha\dfrac{\partial g_1}{\partial \gamma_2} \\ 0 & 0 & 0 \end{pmatrix},$$

where all derivatives are evaluated at the self-confirming equilibrium given in Proposition 4.2. The characteristic equation for the eigenvalues is then

$$\lambda\left(\lambda + \alpha\frac{\partial g_1}{\partial \gamma_1}\right)\left(\lambda + \alpha\frac{\partial g_0}{\partial \gamma_0}\right) = 0,$$

which has roots $\lambda = (0, -\alpha\partial g_1/\partial\gamma_1, -\alpha\partial g_0/\partial\gamma_0)$. Hence, the self-confirming equilibrium is E-stable if and only if

$$-\frac{\alpha\partial g_1}{\partial\gamma_1} < 1 \quad \text{and} \quad -\frac{\alpha\partial g_0}{\partial\gamma_0} < 1.$$

Let us first consider $\partial g_1/\partial\gamma_1$. From (A.3), we have

$$-\alpha\frac{\partial g_1}{\partial \gamma_1} = \frac{\left(1 + \dfrac{1+\lambda\alpha^2}{\beta\gamma_1^2}\right)\alpha g_1}{2\alpha g_1 + \gamma_1 - \dfrac{1+\lambda\alpha^2}{\beta\gamma_1}}, \tag{B.1}$$

and from (A.1), we have

$$\alpha g_1(\gamma) = \rho - \gamma_1.$$

Substituting this into (B.1) and using (30) to simplify, we can write the stability condition as

$$\frac{(1+\beta\rho^2)(\gamma_1 - \rho)}{\gamma_1(1-\beta\rho^2)} < 1,$$

which simplifies to

$$\gamma_1 < \frac{1+\beta\rho^2}{2\beta\rho}.$$

This is satisfied whenever $\rho > 0$ by (A.5).

Finally, turning to $-\alpha\partial g_0/\partial\gamma_0$, we have

$$-\alpha\frac{\partial g_0}{\partial\gamma_0} = \frac{\alpha^2(\lambda+\beta p_2)}{1 - \beta\gamma_1 - \beta(1+\beta)\alpha g_1(\gamma) + \alpha^2(\lambda + \beta p_2)}.$$

Using the facts that $p_2 > 0$ and $\alpha g_1(\gamma) = \rho - \gamma_1$, the condition $-\alpha\partial g_0/\partial\gamma_0 < 1$ will be satisfied if

$$1 - \beta\gamma_1 + \beta(1+\beta)(\gamma_1 - \rho) > 0,$$

which simplifies to

$$1 - \beta\rho + \beta^2(\gamma_1 - \rho) > 0,$$

which is true if $0 < \rho < 1$, given that (A.5) implies $\gamma_1 > \rho$ when $0 < \rho < 1$.

APPENDIX C

ESCAPE DYNAMICS WITH TWO-SIDED LEARNING

Escapes occur from the SCE. Approximate the constant gain learning algorithm (38) by the diffusion

$$d\theta = D_f(\theta^*)(\theta - \theta^*) + \Sigma^{1/2}(\theta^*)dW \qquad \textbf{(C.1)}$$

where $\theta = (\gamma, \delta)$, θ^* is the SCE, $D_f(\theta^*)$ is the Jacobian of the mean dynamics evaluated at the SCE (scaled by the appropriate R matrices), and Σ is the steady state covariance matrix of the innovation vector, μ_{t+1} (again scaled by R). From Williams (2001) and Bogomolova et al. (2005), escapes are determined by the Gramian, G, of (C.1), defined by $D_f \cdot G + G \cdot D_f' + \Sigma = 0$. Specifically, the (local) escape direction is given by the eigenvector of G associated with the largest eigenvalue. [Bogomolova et al. (2005) argue that this can be approximated by the largest eigenvalue of Σ.] In our two-sided learning model, the 5×5 matrix Σ (and therefore G) is singular; it has rank 3. (Notice that at the SCE the bottom two rows of μ are identical, and the second row is $\bar{y}$ times the first.) This means that the escape dynamics are governed by a three-dimensional subspace, spanned by the eigenvectors associated with the nonzero eigenvalues. In particular, δ lies on a manifold determined by γ. Essentially, movements in δ are "costless," because they can be made consistent with the model's mean dynamics by an appropriate change of coordinates. Loosely speaking, this is because the private sector introduces no new shocks to the system.