Paul Krugman

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Are Currency Crises Self-Fulfilling?

1. *Introduction*

In recent years there has been a major revival of interest in the modeling of currency crises. This revival has been driven in large part by events: the series of crises that partially wrecked Europe's Exchange Rate Mechanism (ERM) in 1992–1993, and the Mexican crisis of late 1994 and its aftermath. The new interest has also been driven, however, by the exciting policy conclusions of new models, most of them inspired by the seminal paper of Obstfeld (1994).

What differentiates the new currency-crisis literature from the "classical" literature exemplified by Krugman (1979) and Flood and Garber (1984)? One important difference is a change in the macroeconomic and policy models that are used to describe crisis-prone countries. The old currency-crisis models were essentially seignorage-driven: countries were assumed to have an uncontrollable need to monetize their budget deficits, and to face crisis when this need collided with the attempt to maintain a fixed exchange rate. Obstfeld and his followers have pointed out that this is a very poor description of the position of such recent crisis countries as Britain and Italy in 1992, and is not even a very good description of Mexico in 1994. Instead, the policy dilemmas facing these countries have centered on such issues as real overvaluation, interest rates, and unemployment; rather than facing a sharply defined reserve constraint, the governments that experienced currency crises were trying to make the best of trade-offs among these objectives, with speculators attempting to second-guess government intentions as well as capabilities. Much of the recent theoretical effort has therefore gone into trying to develop more realistic models of crisis-prone economies.

This paper was prepared for the NBER Annual Macroeconomics Conference. I would like to thank Maury Obstfeld and Tim Kehoe, my discussants at that conference, for pointing out the holes in that earlier version.

Most of the recent papers have, however, argued that this change in modeling strategy has consequences that go beyond changing the labeling of the axes or the details of the mechanics of crisis. Rather, they argue that when exchange-rate policy is driven by macroeconomic tradeoffs rather than brute monetary concerns, there is a change in the whole logic of currency crises: instead of being events that are in principle predictable, determined by underlying fundamentals, such crises become in large part the result of self-fulfilling expectations. As Obstfeld (1995) puts it, the "new generation of crisis models suggests that even sustainable pegs may be attacked and even broken"—that is, a fixed exchange rate that could or would have lasted indefinitely in the absence of a speculative attack may collapse simply because financial markets are persuaded, perhaps by otherwise irrelevant information, that the rate will not be sustained.

Some authors have been willing to draw strong policy implications from this conclusion. Most notably, Eichengreen, Rose, and Wyplosz (1995) have argued that the possibility of self-fulfilling crises makes a combination of fixed exchange rates and free capital mobility unworkable; they argue that monetary union and/or capital controls are the only sustainable alternatives to floating. More generally, if we accept the idea that many currency crises are unjustified by fundamentals, there is a strong case for reconsidering the traditional economist's benign attitude toward financial markets: instead of regarding speculators as essentially blameless, mere messengers bringing the bad news, the new models suggest that the George Soroses of the world may be true villains, tearing down structures that might otherwise have stood indefinitely.

These are remarkable conclusions to emerge from no more than a reconsideration of macroeconomic modeling strategy. Can changing the way we represent the government's objective function really make this much difference?

In this paper I want to argue that the answer is no—that the new currency-crisis models, while they have made an important contribution, do not in general imply as radical a rethinking of the logic of crisis as their creators have suggested. More specifically, I will argue that the indeterminacy in the new models does *not* arise from the difference in macroeconomic structure between these models and the "classical" crisis models. Instead, the key change is in the assumptions concerning long-run sustainability. In the classical models, economists envisaged a situation in which underlying fundamentals were persistently deteriorating and focused on the timing of an eventually inevitable collapse. More recent modelers have put on one side the possibility of secular trends in fundamentals; it is this, not the change in the definition of these fundamentals, that makes the timing of speculative attack arbitrary.

We may also argue, albeit less strongly, that in either the classical or the new crisis models the knowledge that fundamentals will or might deteriorate tends to limit the possibilities for multiple equilibria—specifically, to narrow and perhaps eliminate the gap between the necessary and sufficient conditions for speculative attack. Less strongly still, we may argue that large agents of the George Soros type also narrow this gap, tending to provoke crises as soon as the necessary conditions are satisfied.

Finally, this paper argues that the actual currency experience of the 1990s does not make as strong a case for self-fulfilling crises as has been argued by some researchers. In general, it will be very difficult to distinguish between crises that need not have happened and those that were made inevitable by concerns about future viability that seemed reasonable at the time.

The remainder of this paper is in nine sections. Section 2 offers a brief restatement of the "classical" theory of currency crises, in a form intended to stress some similarities with the more recent literature. Section 3 sets out a reduced-form "new" crisis model, intended to represent the large class of such models developed in the last few years. Section 4 examines what happens when this model is applied to an economy experiencing a secular deterioration in its fundamentals. Sections 5 and 6 examine the role of two kinds of uncertainty—uncertainity about the government's determination to defend the currency regime, and uncertainty about future fundamentals. Section 7 explores briefly the potential role of Soroi—large agents who may be able to provoke currency crises for fun and profit. Section 8 reviews recent empirical literature, and asks to what extent the evidence really does indicate an important role for self-fulfilling crises. Section 9 offers a reexamination of the ERM crises of 1992–1993 in light of the models presented in the paper. A final section attempts to summarize the state of play.

2. *The Classical Crisis Model*

The classical model of currency crises may be said to have originated in the work of Salant and Henderson (1978), who showed why an attempt to peg the price of gold using a government-held stock should eventually end in a speculative attack that abruptly wipes out that stock. This analysis was directly adapted by Krugman (1979) to the case of a country using a stock of reserves to peg its exchange rate; some unnecessary

complications in that model were removed, and its results greatly clarified, in later work by Flood and Garber (1984) and many others.

For current purposes, it will be most useful to state a simple monetary model of crises in a way that at least at first makes it seem as if there were multiple equilibria, then see how the standard analysis establishes a unique timing for speculative attack.

Consider, then, a country that is attempting to maintain a fixed exchange rate against the rest of the world. We will make strong "monetary approach" assumptions: both full employment and purchasing-power parity obtain, and the domestic interest rate equals the foreign rate plus expected depreciation. Without loss of whatever generality remains we may take the rest-of-world price level to be stable at 1, and assume the rest-of-world interest rate fixed. The demand for domestic money can therefore be written

$$M = EL(\epsilon), \tag{1}$$

where E is the exchange rate (domestic money for foreign), $L(\cdot)$ the real money demand, and ϵ the expected rate of depreciation. The domestic money supply may be written as the sum of domestic credit and foreign exchange reserves:

$$M = D + R. \tag{2}$$

Finally, we assume that the government is running a budget deficit, which it must cover by expanding domestic credit D. As long as it can, however, the central bank will attempt to peg the exchange rate through unsterilized intervention; when it is no longer able to do so, the continuing expansion of D will lead to an inflation rate (and hence depreciation rate) π.

Suppose, now, that we were to take a snapshot of this economy at a particular point in time, without trying to track its future evolution. We might well convince ourselves that there are in fact multiple equilibria inherent in this situation. Suppose that reserves lies in the range

$$0 < R < M - EL(\pi). \tag{3}$$

Then it might seem that the following is true: if the market does not expect an immediate collapse of the fixed-rate regime, then the expected depreciation rate is zero, and since there are positive reserves, the fixed rate is viable. On the other hand, if the market expects the exchange regime to collapse, with subsequent depreciation at a rate π,

then money demand immediately falls by more than the reserves available, so reserves are exhausted in a sudden speculative attack. It might seem, then, that there is a range of reserve levels—what Cole and Kehoe (1996a, b) call a "crisis zone"—within which speculative attacks can occur with arbitrary timing, and constitute self-fulfilling crises.

This is not, however, the way that such models are usually analyzed. Why? Because the multiplicity of outcomes can be ruled out through a process of backward induction.

Bear in mind that as described, the situation is one in which the central bank is steadily losing reserves. Thus if we imagine the fixed rate avoiding any speculative attack, it will nonetheless eventually collapse all the same. At that point there would be a discrete drop in the demand for money, as the expected depreciation rate rose from 0 to π; since the money supply would not fall (reserves being exhausted), that would mean a step depreciation of the currency.

But such a step depreciation would offer investors the prospect of a forseeable capital gain at (in continuous time) an infinite rate. It would therefore be in their interest to shift out of the currency a bit before reserves would be exhausted—that is, to launch a speculative attack when reserves fell close to but not all the way to zero. Such an attack would, however, force a collapse of the fixed exchange rate at this earlier date—again offering a step depreciation of the currency, inducing investors to attack still earlier. One can work backwards in this fashion, always finding that the speculative attack must occur earlier, until one reaches a level of reserves so high that it would not be exhausted even if investors believed that the exchange regime is about to collapse. This critical level of reserves is defined by

$$R = EL(0) - EL(\pi) = M - EL(\pi). \tag{4}$$

In short, the standard analysis predicts that a currency crisis will occur *as soon as a speculative attack can succeed.* The range of indeterminacy—the range over which an attack would succeed if it occurred, but seemingly need not occur—is eliminated by reasoning backward from the known eventual collapse of the exchange regime.

It is important to realize what is meant here by saying that multiple equilibria are ruled out. The mechanism that might seem to imply self-fulfilling crises is not being questioned: a speculative attack triggers a change in policy that validates that attack. Nor does the classical analysis deny that there is a "crisis zone," a range of reserve levels within which such an attack can take place. The claim is instead that one will not see

countries with fixed exchange rates living for extended periods inside that zone, because a crisis will occur as soon as they enter it.[1]

It should be immediately apparent that the elimination of multiple equilibria in this case has little to do with the way that fundamentals are modeled—with the "monetary approach" character of the model, or the crude representation of the government as a mechanism that pegs the exchange rate until it literally runs out of money. It is, rather, the assumption that the fixed rate is known to be ultimately unsustainable that establishes a unique relationship between fundamentals and the timing of crisis.

With this review of the classical crisis model, let us then turn to the "new" approach.

3. *The "New" Crisis Models*

"New" models of currency crises come in a variety of types, and differ widely in their details. Arguably, however, we may think of the typical model as telling the following story:[2] A government—no longer a simple mechanism like that in the classical model, but rather an agent trying to minimize a loss function—must decide whether or not to defend an exogenously specified exchange rate parity. In making this decision, it takes three concerns into account.

First, there is some reason why, other things equal, the government would like to have an exchange-rate depreciation. This might involve a desire to reduce unemployment when wages are sticky in nominal terms; or it might reflect a desire to reduce the real value of a heavy domestic debt burden. In any case, there is some payoff to depreciation per se.

Second, the cost of remaining with the fixed exchange rate is higher, the greater the rate of depreciation that private agents expect. In practice, this cost normally takes the form of expectations of depreciation leading to higher interest rates, which in turn have adverse effects either on the budget or on the private economy.[3] Regardless of the details, it becomes more expensive *not* to depreciate the more the financial markets are convinced that you will in fact depreciate.

Third, and offsetting these concerns, the government is reluctant to depreciate for some reason—typically, because it has staked its credibility on the maintenance of the current parity, and would pay a political price (or find that inflation–output trade-offs, interest rates, etc. have worsened) if it abandoned its peg.

We may capture all of these concerns with a simple, reduced-form representation in discrete time.[4] Let e be the logarithm of the exchange rate, with e^* the rate that the government would choose if it faced no credibility concerns, $\bar{e}$ the parity to which it has staked its reputation, and ϵ the expected rate of depreciation, $e^E - e$. It is not necessary to assume any particular functional form, but for simplicity let us suppose that the government's loss function takes the form

$$H = [a(e^* - e) + b\epsilon]^2 + R(\Delta e), \tag{5}$$

where $R(\cdot)$ takes on the value 0 if the government does not allow the exchange rate to change, but takes on the value C if it does. Thus C is a fixed "reputation" cost the government will incur if it abandons its parity.

Let us assume that the government can choose the exchange rate (implicitly, we may think of this as involving monetary policy). If the peg is to be abandoned, then the government may as well go to its otherwise preferred exchange rate e^*; once it does so, the market should not expect any further change, so abandoning the peg would eliminate the first two terms in (5). If the government is currently pegging, on the other hand, the market might expect either that it will continue to do so ($e^E = \bar{e}$) or that it will abandon the peg next period ($e^E = e^*$). The decision about whether to retain the peg this period will then depend on the comparison of the loss from staying on the peg with the credibility cost of leaving it; that is, on whether

$$[a(e^* - \bar{e}) + b(e^E - \bar{e})]^2 > C. \tag{6}$$

1. This distinction is crucial in assessing historical experience. If you conclude that Britain would not have dropped out of the ERM in September 1992 had it not been for the speculative attack, this is *not* evidence in favor of self-fulfilling crises—you would say the same thing following a classical currency crisis whose timing was entirely determinate. What you must conclude, rather, is that a similar attack would have driven Britain out even if it had occurred several months earlier, implying that Britain had lived for some length of time within the crisis zone.
2. Some formal models do not quite work this way. Obstfeld (1995) offers an informal exposition that seems to correspond quite well to the description here; but his formal model does not, as explained in footnote 3.
3. Some recent models do not fit this description. For example, in the formal model offered in Obstfeld (1995), *past* expectations of depreciation, as reflected in the predetermined current level of wages, affect the government's decision about whether to devalue; but expectations of *future* depreciation play no role. In this model one cannot use backward induction to tie down the timing of crisis, essentially because nobody has an incentive to look more than one period ahead. Thus the approach taken here does not represent the full range of recent literature.
4. The Appendix offers an illustrative particular model which gives rise to this loss function.

Suppose that the market does not expect a depreciation. Then the second term in (6) will vanish, and the government will want to maintain its peg, fulfilling the market's expectations, as long as

$$[a(e^* - \bar{e})]^2 < C. \tag{7}$$

Suppose on the other hand that the market does expect a depreciation. Then the second term will become positive, and the government will abandon the peg and once again fulfill expectations as long as

$$[(a + b)(e^* - \bar{e})]^2 > C. \tag{8}$$

Clearly, then, we have multiple equilibria as long as

$$[a(e^* - \bar{e})]^2 < C < [(a + b)(e^* - \bar{e})]^2. \tag{9}$$

As long as the economy's parameters put it in that range, either expectations that the exchange regime will survive or expectations that it will collapse will be confirmed by government action.

In the next part I will offer some reasons to question the reasonableness of this result. Even before doing so, however, it may be worthwhile pointing out some limits to the policy relevance of the analysis.

Some discussions of the implications of the new crisis models, notably Eichengreen, Rose, and Wyplosz (1995), seem to blur the line between the proposition that *some* potentially sustainable fixed-rate regimes can be overthrown by speculative attack and the far stronger proposition that *any* fixed-rate regime can be subject to self-fulfilling crisis. It is immediately apparent from (9) that this is not the case: self-fulfilling attacks are possible only over a range of parameters, not for any parameters. Indeed, even this reduced-form representation indicates loosely the conditions for a crisis-proof fixed rate: e.g., a high cost to abandoning the peg (for example, a very strong public commitment), and of course a peg that is not too far from the "right" level (e^* close to $\bar{e}$).

One might argue that the actual evidence shows that fixed rates have collapsed when they were clearly sustainable. As we will see shortly, however, it is substantially harder to make that case than seems to have been appreciated.

First, however, let us try to draw a parallel between this "new crisis" model and the classical crisis model presented in Section 2.

4. *The Effects of Deteriorating Fundamentals*

None of the recent crisis models embodies the element that was crucial in pinning down the timing of crisis in the classical crisis model: a secular deterioration in fundamentals.[5] Yet there is nothing about a more complex and sophisticated representation of the government's decision problem that precludes the possibility that fundamentals change over time, and may do so predictably.

Indeed, it is easy to think of a number of realistic ways in which the fundamentals of countries that have experienced currency crises in recent years have shown secular tendencies toward deterioration. A partial list might include the following (entries are numbered so that they may be referred to later):

(i) Persistent "inertial" inflation at rates greater than trading partners' may make a fixed exchange rate increasingly overvalued, increasing the employment cost of maintaining that parity.
(ii) Even a constant unemployment rate may have growing social costs, as families run down their savings, unemployment benefits are exhausted, and long-term unemployed workers are transformed from employable "insiders" to unemployable "outsiders."
(iii) External debt may accumulate due to large current account deficits, leading to questions about the ability or willingness of the country to honor its obligations to foreign creditors.
(iv) Internal debt may accumulate at an accelerating rate, as interest payments exceed the primary surplus, leading to questions about the solvency of the government.
(v) The political position of the government may approach a terminal condition, as mandatory elections approach or as a parliamentary majority is eroded by resignations, defections, and mortality.

For these and other reasons, it is reasonable to suppose that the parameters in the loss function (5) will predictably shift over time, just as reserves predictably decline in the classical crisis model.

In general, any and all of the parameters might shift; but for current purposes let us assume that what actually shifts is e^*, the exchange rate that the government would choose if it were not concerned with credibility. And for the moment let us assume that e^* has predictable upward trend.

5. Cole and Kehoe (1996a, b) develop an infinite-horizon model of debt crises (without a currency component) in which capital and debt may evolve over time; however, the equilibria they study are all Markov, rather than embodying any secular trend.

Suppose that the fixed exchange rate is ultimately unsustainable—that is, there is a future date T at which it is known that $e^*(T)$ will be sufficiently high that the government would abandon the fixed rate even in the absence of a speculative attack. That is,

$$[a(e^*(T) - \bar{e})]^2 > C. \tag{10}$$

Then consider the previous period. Since investors know that the peg will be abandoned in the next period, they will have an expected exchange rate $e^*(T)$, and the peg will therefore necessarily be abandoned in period $T-1$ if

$$[a(e^*(T-1) - \bar{e}) + b(e^*(T) - \bar{e})]^2 > C. \tag{11}$$

We can work backward in this fashion, and discover that the *latest* possible date for a currency crisis is the first period t for which

$$\{a[e^*(t) - \bar{e}] + b[e^*(t+1) - \bar{e}]\}^2 > C. \tag{12}$$

Finally, suppose that periods are short compared with the trend in e^*, so that $e^*(t+1)$ is close to $e^*(t)$. Then (12) may be approximated by the sufficient criterion for currency crisis

$$\{(a+b)[e^*(t) - \bar{e}]\}^2 > C. \tag{13}$$

Referring back to (9), what we therefore see is that the gap between the necessary and sufficient conditions for currency crisis—between the parameter values for which a crisis *can* happen and those for which it *must* happen—has vanished, and so therefore have the multiple equilibria. Just as in the "classical" crisis models, the knowledge that the fixed rate is ultimately unsustainable means that a speculative attack must occur at the earliest time at which it can succeed.

The recent currency-crisis literature, then, has been wrong in suggesting that the shift from a mechanical seignorage-and-reserve-exhaustion model of crisis to one in which governments minimize a realistic loss function is per se a source of multiple equilibria. As long as there is a secular trend in the fundamentals (defined as fuzzily as one likes) that must eventually make the exchange rate unsustainable, the logic of currency crises becomes a matter of timing, and multiple equilibria disappear as an issue.

One may, however, reasonably argue—for both the new and the classical crisis models—that this logic neglects important uncertainties facing

speculators. These uncertainties may be of two kinds. First, the government's loss function must be a matter of conjecture until it is put to the test, at which point it may turn out that the government is either less or more willing to defend the regime than expected. Second, the assumption that fundamentals inexorably deteriorate is too strong. Surely governments sometimes reverse policy direction, seemingly overvalued currencies begin to look undervalued with the emergence of new export opportunities or declines in world interest rates, or the ability of governments to stay the course turns out to be greater than anyone expected. How do these uncertainties affect the analysis?

5. *Uncertainty about the Loss Function*[6]

During both the European crises of 1992–1993 and the Latin American crises of 1994–1995, individual governments surprised many observers—myself included—who had misjudged the depth of their commitment to fixed rates, in both directions. The speed with which Britain's Chancellor of the Exchequer went from Churchillian rhetoric about defending sterling to proclamations that the devaluation of the pound had him singing in the bath was startling; so was the determination of France to maintain the *franc fort* despite ever worsening unemployment and budget woes. Mexico's unwillingness during 1994 to match monetary policy to the goal of a strong peso was surprising; so was the way that Argentina, despite more than 20% unemployment and a massive banking crisis, held firm to its one-for-one parity between pesos and dollars. But such surprises are themselves unsurprising: the only way for anyone (including the government itself) to be sure about a government's loss function is to put it to the test.

We may crudely represent this kind of uncertainty as follows: As in the previous section, we suppose that fundamentals will predictably deteriorate. However, the fixed cost C that the government perceives itself as facing if it abandons the currency peg is now uncertain. With a probability p it takes on a low value, C_1; with a probability $1-p$ takes on a higher value, C_2.

Using the logic of the preceding section of this paper, it is clear that the currency must be attacked by the time that the fundamental e^* has deteriorated to the point where

$$[(a+b)(e^* - \bar{e})]^2 > C_2. \tag{14}$$

6. This discussion is similar to, but somewhat more careful than, the discussion in Krugman (1979) of the "one-way option" created when it is uncertain how much of its reserves the government is actually willing to commit to defending the exchange rate.

Will it be attacked earlier? That depends on whether (14) is a more or less stringent criterion than the following:

$$[(a + pb)(e^* - \bar{e})]^2 > C_1. \tag{15}$$

It could turn out that any level of the fundamental e^* that satisfies (15) also satisfies (14); this will be true either if the probability p that the government is relatively willing to cave is low, or if the difference between C_1 and C_2 is small. In that case the timing of the speculative attack will be determined by the criterion (14). But if (15) implies a less stringent test than (14)—that is, if it is satisfied for a lower value of e^*—then as soon as fundamentals reach that level there will necessarily be a "probing" speculative attack that tests the government's resolve.

To see why, first consider the situation one period before the period T for which (14) is satisfied. If the fixed rate has survived to that point, it will be known that it collapses in the next period; so the expected rate of depreciation will be

$$\epsilon = e^*(T) - \bar{e}. \tag{16}$$

But if (15) really is a less stringent condition than (14), then if the cost to abandoning the fixed rate really does take on its low value, given this expected rate of depreciation the government will abandon the fixed rate in period $T - 1$.

Now consider the situation in period $T - 2$. Investors know that the government will abandon the parity in $T - 1$ if it has low C; so their expected rate of depreciation is

$$\epsilon = p[e^*(T - 1) - \bar{e}]. \tag{17}$$

This will, however, lead to an abandonment of the parity in $T - 2$ if C is low and (15) is satisfied. We can therefore step back to $T - 3$ and make the same calculation; and so on. We finally reach the conclusion that an attack must occur as soon as (15) is satisfied; at that point the expected rate of depreciation will shoot up to $p(e^* - \bar{e})$.

The attack need not succeed. If the government really does turn out to have a high subjective cost to abandoning the parity, it will demonstrate that by defending the fixed rate despite the need to do so by imposing higher interest rates; once the demonstration has been made, the expectation of devaluation will vanish for a time, until fundamentals deteriorate to the point at which a second, more decisive speculative attack must occur.

Could an attack occur even earlier? No: by construction, if the government has a high C, such an attack will fail even if investors are completely convinced that it will succeed; so even a speculative attack driven by the false belief that the exchange regime must collapse will generate a true expected rate of depreciation of only $p(e^* - \bar{e})$, which again by construction is insufficient to lead to abandonment of the parity, even if C is low.

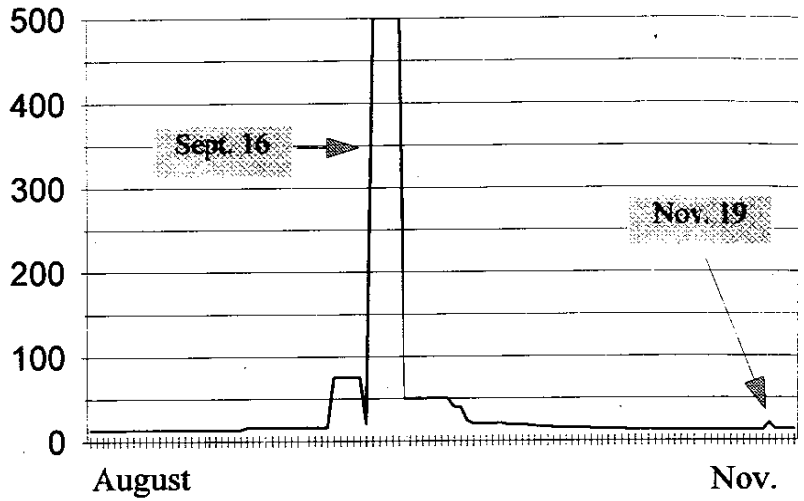
In short, uncertainty about the government's loss function does not in itself generate any indeterminacy about the timing of speculative attacks. Instead, it creates a pattern of "probing" attacks at determinate times that test the government's willingness to defend the currency, then recede if it proves indeed to be willing to pay the price of sustaining the fixed rate. (Notice that the market is not deliberately trying to elicit information about the government's loss function—this behavior is a consequence of individual and indeed atomistic efforts to maximize profits.)

This analysis suggests that one needs to be very careful in drawing loose conclusions from historical episodes of speculative attack, bearing in mind that such episodes themselves elicit information that we have in hindsight but that markets did not have ex ante. On one side, we may look at the collapse of sterling's ERM parity and conclude, as Obstfeld and Rogoff (1995) do, that "the speculative attack on the British pound in September 1992 would certainly have succeeded had it occurred in August." What do we mean by this? Given that we now know that Norman Lamont's rhetoric about defending the pound was largely bluff, we can conclude that if speculators had decided *with certainty* in August 1992 that sterling would drop out of the ERM, that expectation would have been validated; but speculators did not know then what we know now.

Consider, in particular, the contrary example of Sweden, which offers a clear example of the case of probing attacks. Sweden allowed the krona to float on November 19, 1992 in the face of a speculative attack. Looking at that decision, and at the subsequent large depreciation against the DM, one might be tempted to conclude, just as in the case of sterling, that the attack that pushed the krona off its peg would surely have succeeded had it taken place a month or two earlier. Figure 1, however, shows the marginal rate charged by the Swedish central bank—a useful indicator of monetary policy—from August through November 1992. As we can see, in fact there *was* an earlier attack on the krona, in September following the sterling crisis—an attack that failed when the Swedish government proved ready to defend the currency with very high interest rates. Might not the same have happened to an attack on sterling in August?

Conversely, it is tempting to look at speculative attacks that failed–

Figure 1 SWEDISH INTEREST RATES



such as the tequila effect that shook Argentina but in the end did not push the peso off its parity—as evidence of irrational or herding behavior by the markets; but the markets did not know how much the Argentine government was willing to endure to preserve the parity, and such probing speculative attacks may be both rational and determinate in their timing when the government's objectives are uncertain.

Does this mean that uncertainty offers no reason to resurrect the idea of self-fulfilling crises? No; a different kind of uncertainty may once again create a gap between necessary and sufficient conditions for speculative attack.

6. *Uncertain Future Fundamentals*

The problem of modeling currency crisis when the fundamentals evolve according to a random process is not exactly a new one; precisely that issue underlay the literature on the so-called "gold-standard paradox," a subset of the immense literature on target zones (see Buiter and Grilli, 1992; Krugman and Rotemberg, 1992). To the extent that this literature, which made use of the simple reserve-exhaustion model of crisis, found a resolution for this paradox—a very limited resolution at best—it did so by placing restrictions on the *post*crisis regime that restored the presumption that a speculative attack must occur as early as possible. It is difficult

to see how a comparable resolution can be achieved using the new crisis models. The approach described here is far from satisfactory, but it may offer a useful preliminary view. It suggests a plausible answer: that whereas a *certain* eventual unsustainability of a fixed rate eliminates the range of multiple equilibria, a merely *possible* unsustainability simply narrows it.

Let us maintain the basic reduced-form model of the decision whether to remain on a currency peg, as well as the assumption that evolution of the fundamentals over time can be represented by drift in the "otherwise desirable" exchange rate e^*. Now, however, we suppose that e^* evolves randomly. Specifically, we imagine that e^* can only take on one of a number of discrete possible values, indexed by j; let a superscript represent this "step" in the ladder of possible values, so that e^* can take on values e^1, e^2, etc. (It is not necessary to assume that the distance between steps is constant.) And we suppose that at each step there are (possibly step-dependent) probabilities of transition to neighboring steps: from $e^* = e^j$, there is a probability p_j that next period $e^* = e^{j+1}$, a probability $1 - p_j$ that next period $e^* = e^{j-1}$.

Given the uncertain future evolution of e^*, we can no longer use the device of backward induction to find the latest possible point at which the fixed rate must collapse. But we can carry out a corresponding exercise in the space of fundamentals, trying to determine the *least favorable fundamentals under which the fixed rate need not collapse.*

Suppose that there is a level of the fundamentals—call it level J—at which the government would abandon the fixed rate even if there were no speculative attack. That is,

$$[a(e^J - \bar{e})]^2 > C. \tag{18}$$

Now consider the next worst possible level of fundamentals. Is it possible for the exchange rate to remain fixed at that level? The market knows that if fundamentals should worsen, the regime will collapse, so the most favorable expected exchange rate is

$$e^E = p_{J-1}e^J + (1 - p_{J-1})\bar{e}\,. \tag{19}$$

Thus the rate would necessarily collapse at this level of e^* as long as

$$\{a(e^{J-1} - \bar{e}) + b[p_{J-1}(e^J - \bar{e})]\}^2 > C. \tag{20}$$

If this condition is satisfied, one can work back to the next worst level of fundamentals, and so on. The conclusion, then, is that a *sufficient* condi-

tion for currency crisis is that fundamentals have deteriorated to the lowest level j for which

$$[a(e^j - \bar{e}) + bp_j(e^{j+1} - \bar{e})]^2 > C. \tag{21}$$

Once again, we can think of the discrete steps as being small, and approximate this criterion as

$$[(a + pb)(e^* - \bar{e})]^2 > C. \tag{22}$$

The gap between the necessary and sufficient conditions for currency crisis—the range over which self-fulfilling attacks become an issue—may therefore be written as

$$[(a + pb)(e^* - \bar{e})]^2 < C < [(a + b)(e^* - \bar{e})]^2. \tag{23}$$

The size of this range depends on p, which may be interpreted as the probability that fundamentals will worsen in the immediate future. If p is zero—that is, there is no possibility at all that fundamentals will worsen—then (23) reduces to (9). The simple models that have been used to argue for the prevalence of self-fulfilling prophecies may thus be thought of as corresponding to an absence of any concerns about potential future unsustainability. On the other hand, with p equal to one—a wholly predictable deterioration in fundamentals—the model reduces to that of Section 4, with crisis necessarily occurring at the most favorable level of fundamentals at which a speculative attack could succeed, and thus with no range of indeterminacy.

It may be useful to take advantage of the functional form assumed here to rewrite the condition still further. Let e^{max} be the level of fundamentals at which the fixed rate would be abandoned even in the absence of a speculative attack, defined implicitly by

$$[a(e^{max} - \bar{e})]^2 = C. \tag{24}$$

And let e^{min} be the most favorable level of fundamentals for which a speculative attack would in fact succeed:

$$[(a - b)(e^{min} - \bar{e})]^2 = C. \tag{25}$$

Then it is straightforward to show that the worst fundamentals consistent with the absence of a speculative attack are

$$e^* = pe^{min} + (1 - p)e^{max}. \tag{26}$$

Again, if fundamentals are certain to deteriorate, an attack must occur as soon as it can succeed.

Introducing uncertainty in this way cuts both ways if one is debating the relevance of self-fulfilling crises. On one side, it appears that uncertainty—the possibility that the fixed rate might be sustainable forever—allows us to recover the idea that there is a range of parameters for which speculative attack might but need not occur, and in which crises can therefore be self-fulfilling. On the other hand, the possibility of deterioration in the fundamentals narrows that range. Nor is it necessary for private agents to expect that fundamentals will deteriorate: even if the expected direction of change is favorable, the possibility of movement in the other direction limits the range over which a fixed rate can be maintained.

7. *Soroi*

Suppose that due to uncertainty about the future course of fundamentals there exists a substantial range of fundamentals over which currency crisis could but need not occur. This would appear to offer a profit opportunity to a sufficiently large investor. All that such an investor need do is take a short position in assets denominated in the potential crisis currency, and then take the necessary steps to provoke the potential crisis. Nice work if you can get it.

This presumes, of course, that a sufficiently large investor can in fact induce a self-fulfilling crisis. There is a straightforward manner in which this could happen, and then a more diffuse set of possibilities which are hard to pin down.

The relatively straightforward way in which a large investor can provoke a crisis is by the direct effect of his sales. Let us modify the model slightly. Suppose that we make it explicit that the adverse effect of expected depreciation on the government's loss function arises via the domestic interest rate; e.g., we might write

$$H = a(e^* - \bar{e})^2 + b(i - i^*)^2 + C, \tag{27}$$

where i is the domestic interest rate and i^* the foreign rate. And suppose also that assets denominated in domestic currency are regarded by investors as imperfect substitutes for those denominated in foreign currency. Then let A be the net stock of such assets outstanding; the demand for such assets may, crudely, be considered to depend, other things equal, on the difference in expected yields:

$$A = G(i - i^* - \epsilon). \tag{28}$$

Now suppose that a large investor is in a position to sell a significant quantity S of domestic-currency-denominated assets short, raising the effective supply of such assets to $A + S$. Clearly, this will raise i for any given ϵ, and thus raise the cost to the government, other things equal, of maintaining the fixed-rate regime. By the logic of the process described in Section 6 above, this will provoke a crisis earlier—or, to be more precise, at a more favorable level of fundamentals—than would otherwise be the case.

As an empirical matter, one may question the importance of this mechanism. What a large speculator is doing in this case is, in effect, a private sterilized intervention against a currency. Most empirical estimates of the substitutability between assets denominated in different currencies suggest, however, that only a very large sterilized intervention—one beyond the resources even of a George Soros—would be necessary to have a significant impact on the domestic interest rate. Also, governments themselves have the resources to undertake far larger sterilized interventions in defense of their currencies. So one might discount this potential channel for influence of large agents.

Even so, there might still be a powerful role. Consider that the logic of self-fulfilling crises implies that such crises can be set off by "sunspots"—more or less irrelevant events that for whatever reason are taken by private agents as a signal that the currency regime is about to collapse. Clearly, there is an incentive for a large agent first to take a short position in a currency, then manufacture a sunspot, if only he can figure out how.

In fact, this might not be very hard. What is a better sunspot than the very fact that a large agent who is known for doing this sort of thing is selling a currency? The beauty of this scheme is that market participants need not believe that the large agent has better information than they do, nor need they even believe that other participants believe that he does; all that is necessary is that sufficiently many agents believe that sufficiently many other agents believe that sales by George Soros will in fact provoke a crisis.

The possibility of such "internalization" of the potential for crisis means that one may argue loosely that large agents will narrow the gap between necessary and sufficient conditions for crisis. Once the possibility of a self-fulfilling crisis emerges, so does the possibility of a profitable sunspot-manufacture scheme; so large agents will at least sometimes provoke crises at more favorable fundamentals than the worst consistent with maintenance of a fixed-rate regime. Indeed, if one regards such agents as highly effective, then even in the presence of uncertainty the gap between necessary and sufficient conditions for crisis will vanish: as soon as a speculative attack is possible, a large speculator will take a position and then create one.

A final subtlety: as long as market participants believe that large actors will play this role, it may be unnecessary for them actually to do so. As soon as the fundamentals enter the range in which an attack could succeed, investors will reason that the exchange rate is due for imminent collapse through the action of large agents, and they will therefore launch a speculative attack immediately.

This is a very incomplete analysis of the role of large agents; indeed, a complete model would involve many of the same issues that arise in the analysis of corporate takeovers. [In particular, the Grossman–Hart (1981) problem emerges: if everyone knows that George Soros can provoke a crisis, how can he make any profits? Currency noise traders?[7]] However, it does suggest that the role of large traders further limits the likely practical importance of multiple equilibria in the genesis of currency crises.

A further point may be worth making. There is an ancient tradition among government officials in countries subjected to speculative attack of blaming such attacks on nefarious forces—gnomes of Zurich, Anglo-Saxon enemies of Europe, and so on. There is an almost equally ancient tradition among economists of debunking such complaints. If we take the self-fulfilling crisis story seriously, however, we must also concede that the officials have a point: to the extent that sunspots may provoke an otherwise unnecessary crisis, then it makes sense to discourage and possibly even prosecute individuals who deliberately manufacture such sunspots.

8. *Empirical Evidence on the Nature of Crises*

As indicated in the introduction, the new currency-crisis literature was largely inspired by recent events, especially the ERM crises of 1992–1993; more than anything else, the informal observation that these crises could not be easily described as driven by concerns over seignorage and reserve levels led to the emergence of a new style of model. I will turn to the interpretation of the ERM crises in the next section. There is, however, a small, more formal empirical literature which Obstfeld (1995) and Obstfeld and Rogoff (1995) at least interpret as favorable to the case for self-fulfilling crisis.

The most extensive recent empirical investigations of speculative attacks have been carried out by Eichengreen, Rose, and Wyplosz (1995).

7. An interesting start on this kind of analysis has been made by Morris and Shin (1995).

At the risk of oversimplifying their results, one might summarize them as containing three main stylized facts:

1. While many crises are associated with the kinds of evidence that one might expect from "classical" crisis models—large budget deficits, excessive domestic credit creation, and also poor trade performance—many others, and especially the ERM crises, are not.
2. In those crises that are not associated with easily measured policy problems in the runup to crisis, there is generally also an absence of measurable policy deterioration after the crisis; i.e., governments did not ex post (at least given the 8-quarter horizon used in Eichengreen, Rose, and Wyplosz's study) act in a way that appeared to ratify the attack. Again, there was a particular lack of ex post justification in the ERM crises.
3. Finally, those crises that had few obvious explanatory causes were also largely unanticipated by the financial markets—that is, they were not preceded by an increase in interest premia on securities denominated in those countries' currencies. Rose and Svensson (1994) have shown in the particular case of the ERM that there is hardly any visible deterioration in credibility before August 1992.

These are clearly very useful observations. But do they constitute evidence on behalf of the importance of self-fulfilling crises?

Observation 1—that the data do not appear consistent with classical crisis models—suggests that the new crisis models, in which governments are concerned with macroeconomic trade-offs rather than a mechanical reserve constraint, are indeed a better approach for many of the currency crises of recent years. But does this indicate that self-fulfilling crises are important? Only if you believe that the shift from a seignorage-and-reserve account of crisis to a macroeconomics-and-loss-function approach is in itself a reason to believe in multiple equilibria. We have seen, however, that this need not be the case: the reason why multiple equilibria were absent in the classical crisis models was not the monetary character of the crisis but the assumption that fundamentals would predictably deteriorate, and the reason they are present in many of the new models is the tacit assumption that there is no such predictable deterioration. In short, observation 1 tells us what sort of model is appropriate but gives little indication of whether crises are self-fulfilling.

Observation 2—that it is hard to find postcrisis changes in policy that ratify speculative attacks—may perhaps provide some evidence in favor of self-fulfilling crises, in the sense that the opposite finding might have been taken as evidence that the markets were simply anticipating government policy. Once one thinks carefully about this evidence, however, it becomes much less clear how to interpret this negative finding.

Bear in mind that even in those new crisis models that suggest a strong possibility for self-fulfilling crisis, policy variables are supposed to be endogenous—in fact, multiple equilibria arise precisely because a speculative attack may induce a government to change its policy. So the absence of any clear-cut changes in policy following crisis is, strictly speaking, evidence not only against models without multiple equilibria but also against models with them. Or perhaps it would be better to say that this evidence amounts to a demonstration of the weakness of our measures of economic policy: that it simply shows the poor quality of the data.

An alternative interpretation of the evidence in Eichengreen, Rose, and Wyplosz is that what they are measuring is changes in fundamentals—the equivalent of the changes in ϵ^* in the theoretical discussion above. In that case the absence of a clearly defined deterioration in these fundamentals is evidence against any underlying reason for the currency crises. But this interpretation runs into both practical and conceptual difficulties. At a practical level, one may question whether any of the quantitative measures available is a good proxy for the true fundamentals implied by a realistic model of the decision whether to defend a fixed rate: since the decision is essentially political, it is likely to be influenced strongly by the exhaustion of hard-to-measure reserves of public patience and political capital rather than tangible measures like financial reserves. At a conceptual level, one might remark that one main point of the classical crisis models was that an abrupt speculative attack need have no obvious explanatory event: reserves simply needed to fall to a certain critical level, which might be very hard to determine in advance. Similarly, in new crisis models a gradual deterioration in (already hard to measure) fundamentals should eventually push the economy to a critical point at which crisis occurs; one should not expect to be able to spot any break in the trend.

This leaves Observation 3, that the ERM crises (and, to a lesser extent, the Mexican crisis) seem to have come out of a clear blue sky, in the sense that there was little sign of a loss of credibility until shortly before the speculative attacks. It seems to be widely accepted that this supports the idea that the crises were self-fulfilling rather than justified by fundamentals—that the absence of a loss of credibility in financial markets indicates that there was no reason why the currencies in question needed to be attacked. However, on reflection this is not that easy a case to make: if a currency is known to be vulnerable to self-fulfilling speculative attacks, which will lead if successful to a discrete devaluation, then the possibility of such an attack should be reflected in market

expectations even in advance of any actual attack. Even if an attack need not happen, markets should reflect the possibility that it might. In this sense, the absence of any early warning from the financial markets about recent currency crises is as puzzling for advocates of self-fulfilling-crisis stories as for anyone else.

Recently Obstfeld (1995) and Obstfeld and Rogoff (1995) have made an ingenious if sketchy case for regarding the apparent surprise character of recent crises as evidence in favor of multiple-equilibrium stories. The argument runs as follows: sudden speculative attacks have led in a number of cases to large depreciations, which would have offered large profit opportunities to anyone who had foreseen them. Since these opportunities were not reflected in interest premia, the depreciations must have been regarded ex ante as events that were of low probability. And if we assume rational expectations, they must truly have been of low probability given the information available to markets. But where were the large surprise shocks to the underlying economic environment facing or policies carried out by the crisis countries? Obstfeld and Rogoff argue that it is implausible to suppose that there were surprise shocks of sufficient magnitude. If, however, one attributes the crises to sunspots—events that simply happen to trigger self-fulfilling speculative attack—one need not look for large changes in the environment. And one is also free to suppose that such sunspots are rare enough that markets rationally gave them little weight in advance.

It is an ingenious argument, but how convincing is it? Notice that it relies on two ancillary assumptions beyond the self-fulfilling crisis models themselves: the assertion that sunspots that trigger crises are rare, and the assumption of rational expectations on the part of the market. It is unclear why the former should be the case (especially if we bear in mind the discussion of large agents and their incentives, above); it is also true that an overwhelming array of direct evidence suggests that foreign-exchange markets do *not* make use of all available information. The apparent failure of the markets to assign any substantial probability to either the ERM or Mexican crises is indeed a puzzle, as discussed at greater length below; but it is far from clear that a low-probability sunspot story is the right way to resolve this puzzle.

The other obvious point to make is that both the ERM crises and the Mexican crisis were preceded by surprise *political* developments that came as a severe jolt to financial markets, not so much because of their direct impact as because of the revelation that the political basis for the currency regime was less solid than they had imagined. The era of crisis in the ERM began with two shocking referendum results on Maastricht: the initial rejection by Danish voters, and the paper-thin victory in France's referendum. These votes revealed, to most observers' great surprise, that the enthusiasm of Europe's policy elite for EMU was not shared by the broader public; they may thus be regarded as "large" events despite the fact that Danish EMU advocates were able to call for a replay and French advocates technically won. In Mexico, the Chiapas uprising and Colosio's assassination revealed a troubled political scene that, once again, was news to the financial markets (even if they should have known better).

The available evidence, then, does not establish an overwhelmingly compelling case for the importance of self-fulfilling expectations in currency crises.

9. *The ERM Crises of 1992–1993*

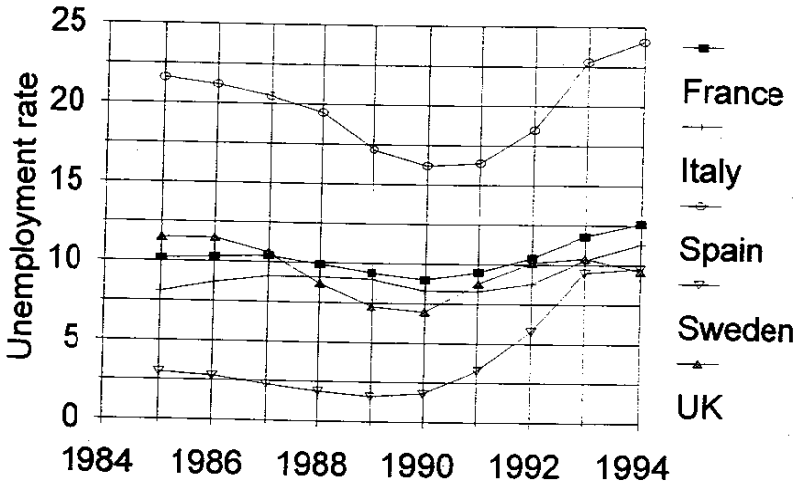
There is no obvious way to test directly whether any particular currency crisis was a necessary event given expectations about fundamentals, or simply a self-fulfilling event triggered by a sunspot. It is possible, however, to ask whether at the time of a crisis the crisis country was experiencing a secular deterioration in fundamentals which, like the gradual erosion of reserves in the runup to "classical" crises, could have been pushing it toward a critical point. If not—if it is hard to see any reason why markets might have concluded that the exchange regime was ultimately unsustainable—then one may be strongly inclined to turn to self-fulfilling-crisis stories. On the other hand, if there is a clearly visible deteriorating trend, the advocate of self-fulfilling-crisis models is placed in the much weaker position of arguing that even though there is an explanation of the crisis in terms of fundamentals, it is quantitatively insufficient.

In this light, let us consider the evidence on the ERM crises of 1992–1993, making use of the checklist of possible types of secular deterioration given in Section 4 above. I focus on five countries whose currencies were attacked: France, Italy, Spain, Sweden, and the United Kingdom.

Consider first the underlying macroeconomic situation of these economies, as measured by four indicators: unemployment, output gaps, inflation, and debt. Some relevant data on each are shown in Figures 2–5. These data point strongly to a simple conclusion: all five economies were, by 1992–1993, in a situation where a standard macroeconomic diagnosis would prescribe monetary expansion—a monetary expansion that was blocked by the ERM. Thus all five economies had an evident incentive to abandon their parities.

Figure 2 illustrates the obvious point that the European recovery of the second half of the 1990s had, by the time of the ERM crises, turned into a

Figure 2 UNEMPLOYMENT RATES



severe and deepening recession. The rise in unemployment had been particularly severe in Sweden and the United Kingdom, least visible in Italy.

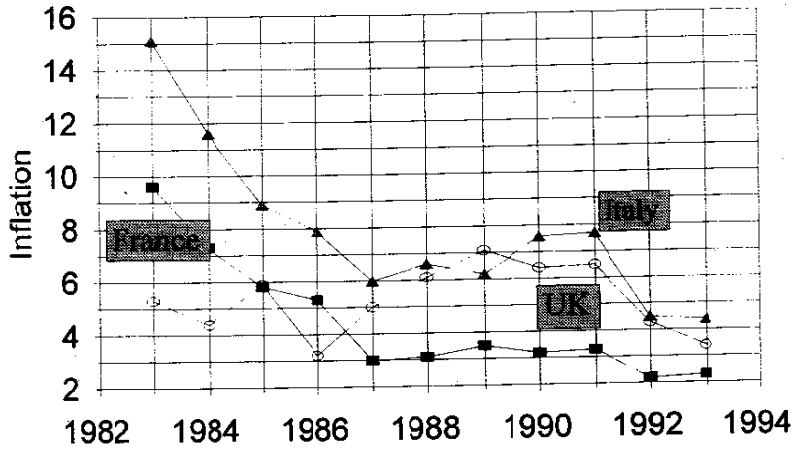
It might be objected that many European countries have shown secular upward trends in unemployment, so that the level of unemployment gives only weak evidence about the scope for monetary expansion. However, standard estimates of output gaps—the difference between the level of output and that consistent with a stable rate of inflation—show even more clearly the deterioration in the early 1990s. Figure 3 shows the European Commission's estimates (European Commission 1995), which are similar to those of other institutions, including the OECD and the IMF.[8]

Consistent with the view that output in the early 1990s had fallen well below its natural rate, Figure 4 shows that after accelerating in the late 1980s in the countries in question, the inflation rate (as measured by the GDP deflator) was both falling and at already quite low levels.

How should we think of the situation implied by these observations? Suppose that your view of the macroeconomy is a textbook natural-rate-plus-adaptive-expectations model—that is, a model in which the inflation rate accelerates when unemployment is below the NAIRU, decelerates

8. The EC estimates are based not on an attempt to estimate a Phillips curve, but on a trend-fitting technique. However, since the actual level of output tends in any case to fluctuate around its "natural" level, the results are similar.

Figure 3 OUTPUT GAPS



when unemployment is above the NAIRU. Then both the comparison with trend output and the falling inflation rates in European countries would be clear indicators that the recession had pushed unemployment rates well above the NAIRU, while the combination of low inflation and high unemployment meant that governments might reasonably feel that reducing unemployment was more urgent than driving inflation still lower. In short, the European economies—other than Germany—were in

Figure 4 INFLATION RATES

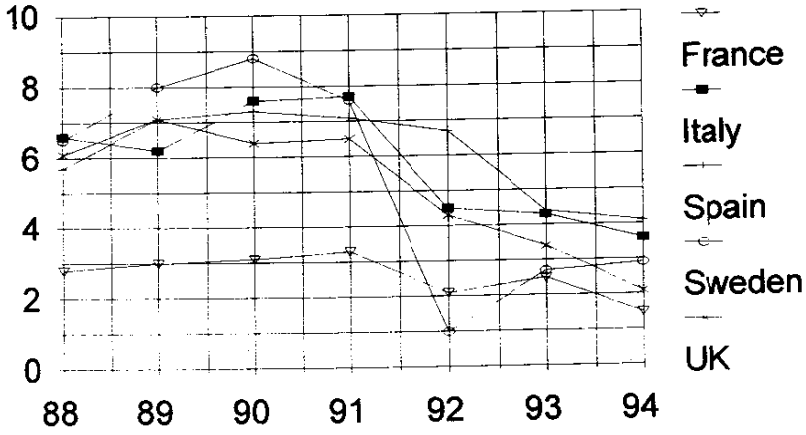
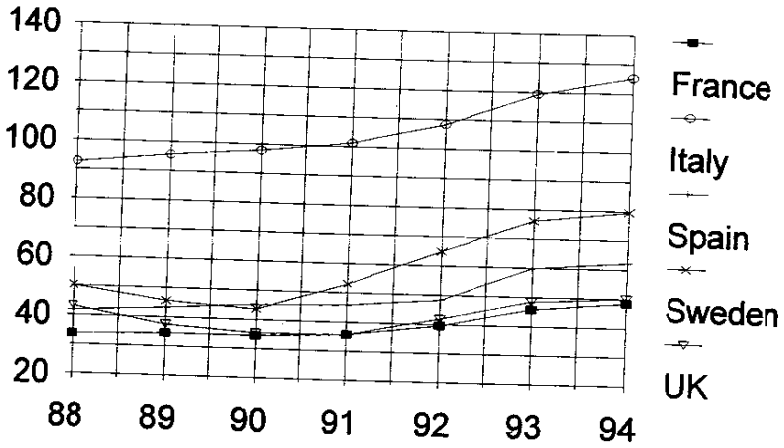


Figure 5 DEBT/GDP RATIOS



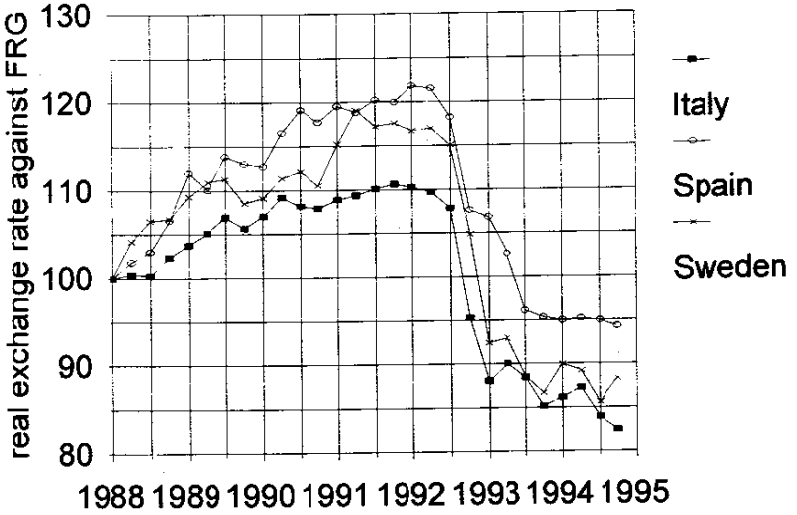
a situation in which the textbook policy recommendation would be an expansionary monetary policy. Unfortunately, a commitment to an exchange-rate mechanism in which Germany acted as de facto key currency country left other European countries no room for independent monetary policy.

The case for a monetary expansion frustrated by the commitment to the ERM is reinforced by the debt situation illustrated in Figure 5, which shows the debt/GDP ratio. The marked deterioration in several countries, especially Sweden and Italy, meant both that fiscal expansion as an alternative to monetary policy was out of the question, and that a monetary expansion—which would have helped reduce outlays on unemployment benefits and increased tax revenues—was that much more attractive.

In short, we can easily make the case that for all five countries e^* was substantially larger than $\bar{e}$—that in the absence of the ERM commitment all five countries would have chosen more expansionary monetary policies, leading to a depreciation of their currencies against the Deutsche mark.

Now let us turn to the question of whether e^* was predictably deteriorating. The simplest form of persistent deterioration in fundamentals is that in which overvaluation grows over time via "inertial" inflation, discussed in Section 4 under the heading (i). Here there is some diversity among the European countries. Figure 6 shows the real exchange rates of Italy, Sweden, and Spain against Germany from the beginning of 1988 to the end of 1994; these three countries continued to experience inflation at more rapid rates than Germany despite being pegged to or (in the case of Sweden) "shadowing" the ecu, and thus became increasingly overvalued up to their abandonment of the parity in late 1992.

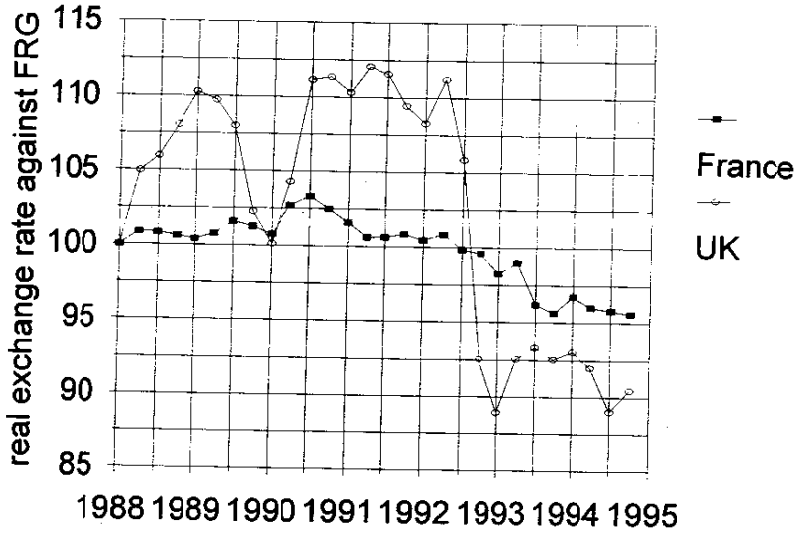
Figure 6 REAL EXCHANGE RATES: DEVALUING NATIONS



The situations of Britain and France, illustrated in Figure 7, were more complicated. The United Kingdom entered the ERM late, and its entry followed a substantial nominal and real appreciation. There was little further real appreciation, but a widespread belief among economists and businessmen that the entry had taken place at too high an exchange rate. France showed little change in its real exchange rate vis-à-vis Germany.

It might seem from this indicator that in the case of France and Britain, while there was a strong incentive to adopt a more expansionary monetary policy and hence drop out of the ERM, there was no obvious reason why that incentive should grow stronger over time—i.e., no secular deterioration in e^*. To many economists at the time, however, it seemed that the pressures for devaluation were growing despite the absence of ongoing real appreciation. After all, unemployment rates and output gaps were rising in both countries; even absent real appreciation, didn't this mean that the gap between e^* and the ERM parity was growing? If one takes standard models of international macroeconomics seriously, the ultimate test of whether a currency is overvalued depends not on intermediate indicators like real exchange rates, but on actual macroeconomic

Figure 7 REAL EXCHANGE RATES: FRANCE vs. UK



performance—whatever PPP calculations say, exchange rates must ultimately be evaluated on a PPE[9] basis.

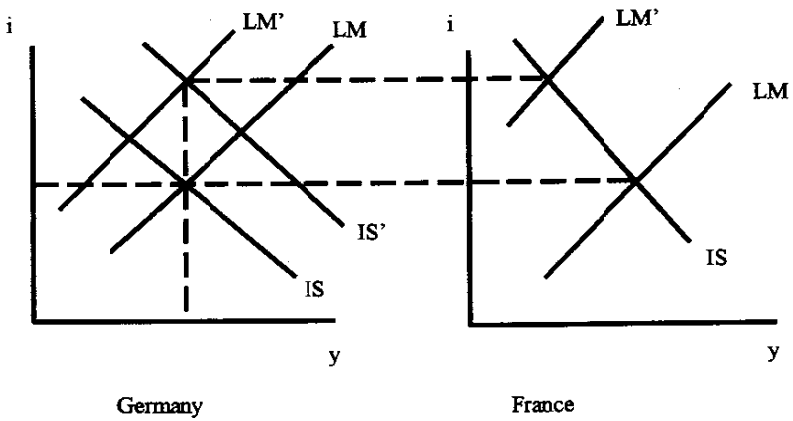
Moreover, there was a particular reason—the interaction between German reunification and the status of the Deutsche mark as the de facto key currency—why many observers believed that there were growing strains on the ERM, leading a number of economists to predict an ERM breakup well in advance of the actual events (see, for example, Krugman, 1990).

This is a familiar story, but it is worth repeating briefly here for the light it sheds on the crisis. Figure 8, drawn from Krugman and Obstfeld (1994),[10] illustrates the standard argument. It shows IS–LM diagrams for two countries: a key currency country ("Germany") and a second country ("France") that has committed itself to using monetary policy to peg its exchange rate. If the exchange-rate peg is fully credible, interest rates must be equal in the two countries.

The scenario then runs as follows: Germany, deciding to finance the costs of reunification with debt rather than current taxes, engages in a fiscal expansion; its IS curve shifts out to IS'. In order to avert any inflationary pressures, however, the Bundesbank offsets this expansion

9. The proof of the pudding is in the eating.
10. Obstfeld's half.

Figure 8 THE LOGIC OF THE ERM CRISIS



with a tight monetary policy, shifting LM in to LM' and leaving output unchanged.

Faced with the resulting rise in the German interest rate, France must match it in order to maintain the currency parity. It therefore is obliged to follow the Bundesbank with its own tight money policy. Since this is not an offset to a fiscal expansion, however, the result is a decline in output, warranted not by the domestic macroeconomic situation but only by the need to maintain parity with the mark.

This is both a crude and a mechanical representation of the economic forces involved, but it nonetheless makes two useful points. First, the fiscal shock from German reunification created a strain on the ERM—a motive for European nations other than Germany to defect from the mechanism—that had not been there before: in 1992 there was conflict between the monetary policy that seemed appropriate for Germany and that which seemed appropriate for other European nations, in a way that had not been the case earlier. Second, the analysis points to the irrelevance of several indicators that commentators have used to argue that France in particular was not a reasonable target for speculative attack. It has been pointed out that in 1992–1993 France had a lower inflation rate than Germany, a smaller budget deficit, and a current-account surplus compared with Germany's deficit; surely, argue some commentators, this means that the franc should have been in a strong rather than a weak position. And they therefore argue that the franc's woes demonstrate that even a country with no fundamental problems can be subjected to a devastating speculative attack. But if

one takes the scenario in Figure 8, and imagines that France and Germany start from identical macroeconomic positions—the same inflation rate, the same budget deficit, and the same current-account balance—one would expect to see France start to look better on all three: a lower budget deficit because it has not had the fiscal expansion, a more positive current account because the depressed state of the economy reduces imports, and over time a lower inflation rate because of the output gap. Nonetheless, in this story France has an incentive to abandon its ERM parity in order to pursue a more expansionary monetary policy; the indicators often cited in support of the idea of a structurally strong franc are irrelevant.

Could it be said that these incentives to depreciate were increasing predictably over time? Again, it is not hard to make a case. First, over the course of 1991–1992 estimates of the cost of German reunification—and hence of the size of the shock illustrated by Figure 8—were rising steadily. Second, as pointed out in Section 4, the political and social strains of a given output gap tend to mount over time. Third, we may once again point to the debt problems, which constituted a visible source of growing pressure (and continue to do so, as recent events in France demonstrate).

On the basis of all of these indicators, then, it is hard to see on what basis one would use the ERM crackup as evidence for self-fulfilling crises. Fundamentals relevant to the willingness of governments to continue pegging their currencies had clearly worsened, and showed every sign of continuing to worsen. These trends caused many economists to forecast a crisis—correctly.

The only argument that one might make on behalf of a self-fulfilling-crisis story would be one that relies heavily on the absence of early warning signs in the financial markets. Interest differentials against crisis countries did not begin to widen until summer 1992. As described above, this observation has been interpreted by Obstfeld (1995) and by Obstfeld and Rogoff (1995) to mean that the attacks must have been low-probability events, instigated by sunspots. The failure of the markets to signal any risks ahead is indeed puzzling. However, consider how the Obstfeld–Rogoff argument stands in the light of the evidence above. We must argue that although there was a substantial deterioration in the fundamentals, which led many economists to forecast a crisis—and *although these forecasts were right*—nonetheless the failure of the markets to anticipate the crisis must be taken as evidence that this crisis was not justified by the fundamentals, and instead was a self-fulfilling event that occurred out of the blue.

What alternative explanation can we offer? It is hard to avoid the suspicion that financial markets were simply myopic in the runup to both the ERM and the Mexican crises. Unfortunately, this conclusion wreaks havoc with all of the currently popular models: both the "classical" view that crises occur as soon as they can, because of forward-looking markets, and the "self-fulfilling" view that crises can occur randomly, because rational investors know that speculative attacks will be validated.

10. Concluding Remarks

Over the last two years international economists have given remarkably serious credence to a view that, if correct, might greatly change our view about the conduct of both macroeconomic policy and financial-market regulation in open economies. This view, grounded in new models of speculative attack, holds that such attacks on fixed exchange rates are not, as has previously been thought, responses to underlying fundamental weaknesses of the currency regime. Rather, they are self-fulfilling events that can undermine otherwise sustainable regimes; some economists seem even to believe that no fixed rate is safe from such attacks.

In this paper I have tried to throw some (but not too much) cold water on this new view. One part of the new view—that governments should be thought of as trading off macroeconomic objectives against credibility, rather than mechanically pursuing credit creation until reserves run out—is surely correct. The new literature goes too far, however, in supposing that this change in the underlying macro and policy model is in itself a necessary reason to believe in multiple equilibria and self-fulfilling crises. A predictable secular deterioration in fundamentals—which was a basic assumption in the old literature—will eliminate the gap between necessary and sufficient conditions for speculative attack in many of the new models as well. Uncertainty of the right kind can restore some indeterminacy in the timing of speculative attacks (in both old and new models), but it can also create a pattern of "probing" attacks that might create a false impression of multiple equilibria. And large agents à la George Soros may act to narrow the range of indeterminacy.

An informal review of the available empirical evidence also casts doubt on the case for self-fulfilling speculative attack. In particular, there seem to have been very good reasons why speculators might have attacked the European countries they did in 1992–1993. It is puzzling that markets did not seem concerned about the possibility of such attacks until very late, especially since many economic analysts had warned about them well in advance; but this lack of early warning can be made into evidence for self-fulfilling-crisis models only through a fairly convoluted and indirect argument.

In sum, we should not take the analysis of self-fulfilling speculative attack too seriously, at least not yet. For the time being it is best to assume that most countries achieve currency crisis the old-fashioned way: they earn it.

Appendix. Deriving the Government's Loss Function from a Simple Macro Model

This paper analyzes the currency-crisis issue in terms of a reduced-form government loss function; the reason for doing so is that the logic of the analysis is largely independent of the details of the macro model. And given the inevitable divisions of opinion about macro modeling strategy, it seems a good idea to put those details aside, so that the main points of the analysis do not get caught up in contentious but orthogonal issues. However, it may also be useful to show how one particular model can give rise to the assumed loss function.

Consider, then, a Mundell–Fleming-type open-economy macro model with sticky prices. (In this model these prices will be treated as a "fundamental"—an assumption that will be reasonable in a medium-term model with substantial inertial inflation. Such a model, it may be argued, is reasonable for thinking about the ERM crises, although not in all cases.) In such models, output is demand-determined; we can linearize the model to write output as a function of the real exchange rate (which determines the competitiveness of the country's goods) and the real interest rate:

$$y = \alpha + \beta(e + p^* - p) - \gamma(i - \pi), \tag{29}$$

where p^*, p are the logs of the foreign and domestic price levels, and π is the expected rate of inflation.

We may also introduce a money demand equation; as this will play no role in the analysis, it can be left generally stated as

$$m - p = L(y,i). \tag{30}$$

The economy is assumed open to capital movement, with equalization of expected returns; thus

$$i = i^* + \epsilon \tag{31}$$

with i^* the foreign interest rate and ϵ the expected rate of depreciation. Finally, we assume that the government's underlying loss function may be stated in terms of the deviation of output from a desired level:

$$H = (y - \bar{y})^2. \tag{32}$$

We may now define the "fundamental" e^* as the (log) exchange rate that would leave output equal to its target level in the absence of any expected depreciation—that is, we define e^* implicitly so that

$$\bar{y} = \alpha + \beta(e^* + p - p^*) - \gamma(i^* - \pi), \tag{33}$$

implying

$$e^* = \frac{1}{\beta}[\bar{y} - \alpha + \beta(p - p^*) + \gamma(i^* - \pi)], \tag{34}$$

which in turn lets us write

$$y - \bar{y} = -\beta(e^* - e) - \gamma\epsilon, \tag{35}$$

leading to the loss function

$$H = [\beta(e^* - e) + \gamma\epsilon)]^2. \tag{36}$$

The logic here is, of course, very simple: output is depressed below its target level both by overvaluation of the exchange rate and by expectations of depreciation, which raise domestic interest rates.

REFERENCES

Buiter, W., and V. Grilli. (1992). Anomalous speculative attacks on fixed exchange rate regimes: Possible resolutions of the "gold standard paradox." In *Exchange Rate Targets and Currency Bands*, P. Krugman and M. Miller (eds.). Cambridge: Cambridge University Press.

Calvo, G. (1995). Varieties of capital-market crises. University of Maryland. Working Paper.

Chen, Z. (1995). Speculative market structure and the collapse of an exchange rate mechanism. London: Centre for Economic Policy Research. Discussion Paper 1164.

Cole, H., and T. Kehoe. (1996a). A self-fulfilling model of Mexico's 1994–5 debt crisis. Federal Reserve Bank of Minneapolis. Staff Report 210.

———, and ———. (1996b). Self-fulfilling debt crises. Federal Reserve Bank of Minneapolis. Staff Report 211.

Eichengreen, B., A. Rose, and C. Wyplosz. (1995). Exchange market mayhem: The antecedents and aftermath of speculative attacks. *Economic Policy* 21:249–312.

Flood, R., and P. Garber. (1994). Collapsing exchange-rate regimes: Some linear examples. *Journal of International Economics* 17:1–13.

Grossman, S., and O. Hart. (1981). The allocational role of takeover bids in situations of asymmetric information. *Journal of Finance* 36:253–270.
Krugman, P. (1979). A model of balance-of-payments crises. *Journal of Money, Credit, and Banking* 11:311–325.
———. (1990). A looming European recession? *US News and World Report*, December 17, p. 73.
———, and Obstfeld, M. (1994). *International Economics: Theory and Policy.* New York: Harper Collins.
———, and J. Rotemberg. (1992). Speculative attacks on target zones. In *Target Zones and Currency Bands*, P. Krugman and M. Miller (eds.). Cambridge: Cambridge University Press.
Morris, S., and H. Shin. (1995). Informational events that trigger currency attacks. Federal Reserve Bank of Philadelphia. Working Paper.
Obstfeld, M. (1994). The logic of currency crises. *Cahiers Economiques et Monetaires* (Bank of France) 43:189–213.
———. (1996). Models of currency crises with self-fulfilling features. *European Economic Review* 40:1037–1048.
———, and K. Rogoff. (1995). The mirage of fixed exchange rates. *Journal of Economic Perspectives* 9:73–96.
Rose, A., and L. Svensson. (1994). European exchange rate credibility before the fall. *European Economic Review* 38:1185–1216.
Salant, S., and D. Henderson. (1978). Market anticipation of government policy and the price of gold. *Journal of Political Economy* 86:627–648.

Comments

TIMOTHY J. KEHOE
University of Minnesota and Federal Reserve Bank of Minneapolis

1. *Introduction*

I have always found Paul Krugman's papers to be thoughtful and provocative, and this paper is no exception. It deals with an important and controversial question: Which of two sets of theories better explain current account crises—the classical theories in which such crises are determined by fundamentals, or the new theories in which, although the possibility of a crisis may be determined by fundamentals, the crisis

I would like to thank, without implicating, Caroline Betts, Chari, Hal Cole, Patrick Kehoe, and Ellen McGrattan for helpful discussions. The research reported on here has been supported by a grant from the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-94-1-0461. The U.S. government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation thereon. The views expressed herein are those of the author and not necessarily those of the Air Force Office of Scientific Research or the U.S. government. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

itself is triggered by what journalists and finance ministers call the herd behavior of investors and economic theorists call, for want of a better term, sunspots? The first set of theories produces crises that are, in the absence of large shocks to the fundamentals, predictable. A Monday-morning quarterback can explain exactly why the crisis should have been foreseen. The second set of theories produces crises with a more arbitrary character. Although we can see the role of fundamentals in determining the conditions that allow the crises to occur, we can also imagine a different outcome.

Paul definitely favors the first set of theories, and not surprisingly—Krugman (1979) was one of the seminal papers in the development of these theories. As economists, we should all favor these sorts of theories *a priori:* ideally, economic fundamentals should pin down outcomes. Recent events, however, especially those in Mexico in 1994 and early 1995, have pushed me in the direction of the second set of theories (see Cole and Kehoe, 1995).

Although Paul's argument that, reinterpreted correctly, the classical theories can still explain the recent current account crises in Europe and Mexico did not convince me, I learned a lot from reading his paper. The next section briefly lays out what I thought to be the most important contributions of the paper. The third section critiques Paul's theory and suggests an alternative in which the economic actors recognize the dynamic nature of the model. The fourth, and final, section argues that the 1994–1995 Mexican crisis had an arbitrary character that is better explained using the second set of theories.

2. *Contributions of the Paper*

In discussing the new crisis theories that have followed the work of Obstfeld (1994), Paul distinguishes between the modeling of endogenous policy and the possibility for multiple equilibria in the models. The decision to devalue is made by a government that acts to maximize welfare in the domestic economy but cannot commit to its future actions. The government therefore faces a time-consistency problem in the sense of Kydland and Prescott (1977). As Barro and Gordon (1983) have stressed, in this sort of environment the expectations of private agents about government actions have an important feedback in determining what those actions should be.

As Paul points out, in a model with endogenous government policy, any economic variable can be a fundamental in terms of explaining a devaluation if we can imagine that variable in the government's objective function. As Paul's discussion of the European Exchange Rate Mecha-