Policy Credibility and the Design of Central Banks

ROBERTO CHANG
The author is a research officer in the macropolicy section of the Atlanta Fed's research department. *He thanks Jerry Dwyer, Adam Posen, Will Roberds, and Mary Rosenbaum for useful suggestions.*

In recent years the practice of central banking around the world has been profoundly affected by two trends. The first is toward granting central banks greater independence vis-à-vis other branches of their governments. This trend is clearly expressed in the British government's May 1997 move granting the Bank of England the power to set short-term interest rates. It is also evident in the current European Union's plan for a single currency: the 1992 Treaty of Maastricht prescribes the creation of a monetary authority, the European System of Central Banks (ESCB), that would be formally independent of any other European government or institution.[1] In addition, many Latin American countries, including Mexico, Argentina, Chile, and Peru, have enhanced the independence of their central banks in the context of broad structural reforms. South Africa's postapartheid government also agreed to an independent monetary authority.[2]

The second trend influencing the nature of central banking is for countries to formally state that a central bank's sole objective should be to ensure price stability. New Zealand, for example, in its Reserve Bank Act of 1989, stated that the Bank's monetary policy should be "directed to the economic objective of achieving and maintaining stability in the general level of prices."[3] Likewise, Article 105 of the Maastricht Treaty establishes that "the primary objective of the ESCB shall be to maintain price stability." The Bank of Canada and some other central banks are now bound to follow formal inflation targets. In many other countries there is considerable debate about whether their monetary policy should be exclusively geared toward attaining zero inflation.[4]

These two trends have an underlying unity: they can be seen as social responses to a more fundamental problem of central bank credibility called the time inconsistency of monetary policy. To aid in understanding this connection, this article discusses the nature of the time inconsistency problem and its economic implications.

The theory of time inconsistency stresses that monetary authorities are often tempted to promise low inflation now and to try to surprise the public with unexpectedly higher inflation later. However, such promises will not be believed because economic agents, understanding the authorities' incentives, realize that the promises will not be honored. Instead, economically plausible outcomes have the property that monetary authorities are not able to systematically surprise the

public. As this discussion will show, this property implies that the monetary authority cannot profit from reneging on its announcements. In fact, it can only lose by doing so: expected and realized inflation will often be higher than if the monetary authorities had made a binding promise. This consequence is known as inflation bias.

This article explains how the creation of some institutions can be interpreted as social responses to time inconsistency. A society may try to ameliorate inflation bias by providing appropriate incentives for its monetary authorities to adhere to promises; institutional arrangements may be designed to reduce the gains to the authorities from creating unexpected inflation. One approach is to structure the compensation of central bankers so as to punish them if inflation is outside some target range, as in New Zealand. Alternatively, a society may try to constrain the policy instruments available to the monetary authorities in order to make engineering inflation surprises more difficult. A country's commitment to fix its exchange rate can be understood in this way. For either approach to work, it is necessary that the monetary authorities be insulated from the rest of the government. Hence central bank independence emerges as a necessary condition for institutional solutions to time inconsistency.

Further theoretical analyses imply that such institutional mechanisms may not be necessary, however. In particular, because monetary authorities are typically engaged in a long-term relationship with the public, they can develop a reputation for honoring commitments. The fear of losing a reputation for future "honesty" is an important incentive that may deter a central bank from "cheating" today. Recent studies have shown that this incentive may be powerful enough to make socially optimal outcomes attainable, even in the absence of any institutional constraints.

Institutional approaches and reputational concerns are both plausible solutions to the time inconsistency problem, and both have weaknesses according to existing theory. To aid in understanding their relative merits, this article discusses related empirical work. Empirical studies have largely focused on testing the hypothesis that the central banks that are more independent deliver lower inflation. Evidence favoring that hypothesis has been analyzed in several studies focusing on developed countries. However, it will be seen that the relationship between central bank independence and inflation seems fragile, and it does not hold for less developed countries.

Although the empirical findings provide little support that central bank independence helps lower inflation, it is too early to discard existing theory. According to the theory, central bank independence is only one aspect of institutional solutions to inflation bias. It cannot by itself eliminate inflation bias, so its emergence will not necessarily yield lower inflation. In addition, reputation-based approaches imply that inflation bias may be addressed by noninstitutional means; hence, low inflation need not require central bank independence. Both arguments imply that there need not be a negative relation between central bank independence and inflation even if current theory is valid.

A trend influencing the nature of central banking is for countries to formally state that a central bank's sole objective should be to ensure price stability.

An Economic Theory of Credibility

Although the role of credibility in monetary policy has been recognized for a very long time, modern research on credibility started only in the late 1970s with the publication of seminal papers by Calvo (1978) and Kydland and Prescott (1977). These two papers showed that the then-novel hypothesis of rational expectations had profound implications for the credibility of macroeconomic policy in general and monetary policy in particular. Before focusing on these implications, it may be helpful to illustrate the basic nature of Calvo's and Kydland and Prescott's ideas with a simple example.

The example is about a fictional father, Federico, and his adolescent son, Pablo, at the start of some week. Pablo is at that age when he dislikes hard work and loves to be extravagant. Federico wants to teach Pablo the value of hard work, of course, and to that end he has convinced the neighbors to let Pablo mow their lawn for money. Federico's problem is that he cannot force Pablo to do the job. Instead, Pablo must be induced to mow the lawn, and the way to convince him is to allow him to get a tattoo in exchange for his effort. Federico would like to prevent Pablo from being tattooed, although this objective is not as important to him as inducing Pablo to mow the neighbors' lawn. Federico would rather have Pablo mow the neighbors' lawn and use the corresponding payment to

1. *See specifically Article 107 of the Maastricht Treaty.*
2. *See "Role Shifts for Central Bankers," in the* New York Times*, November 15, 1994, sec. D.*
3. *Section 8, Reserve Bank of New Zealand Act of 1989, quoted in Walsh (1995b).*
4. *For debate about U.S. policy, see, for instance, "A Matter of Demeanor,"* Wall Street Journal, *May 20, 1994, sec. A, and "Time for an Economic Summit,"* Wall Street Journal, *September 28, 1994, sec. A.*

pay for, say, a good book. The prospect of reading a good book is not enough to induce Pablo to do the lawn, though.

To make the example interesting, assume that the neighbors, mindful of Federico's dilemma, will give the money to Federico and not Pablo. Finally, let us push the fictional nature of the example and assume that it is the only interaction that Federico and Pablo will have.

What is the likely outcome of this father-son example? Federico cannot convince Pablo to mow the neighbors' lawn without promising him a tattoo. It seems that it should be enough for Federico to tell Pablo, "If you mow the neighbors' lawn, you will be allowed to use their payment for whatever you want." If Pablo believes this offer and Federico honors his word, then Pablo will mow the lawn, get the money, and run to the tattoo shop.

A society may try to constrain the policy instruments available to the monetary authorities in order to make engineering inflation surprises more difficult.

However, after Pablo mows the lawn, it is no longer in Federico's interest to allow Pablo to be tattooed. Hence, instead of giving the lawn-mowing money to Pablo, Federico will go to a bookstore and buy Pablo a good book. Then he will just tell Pablo, "Sorry, Son, a tattoo will leave an indelible mark on your body, and I cannot let you have one. Here is a good book for your effort." By breaking his promise in this fashion, Federico would have obtained his most preferred outcome: he will have induced Pablo to mow the neighbors' lawn and also prevented him from being tattooed.

The paradox is that Federico's ability to renege on his promise and surprise Pablo turns out to be counterproductive. If Pablo is intelligent enough to understand his father's decisions, he will not believe Federico's promise and, consequently, he will not mow the neighbors' lawn. Federico's promise is "incredible."

Importantly, Federico ends up worse off than if he could bind himself to honor his word. If he could, he would be able to convince Pablo to mow the lawn. Although he would have to allow Pablo to get a tattoo in order to achieve this goal, Federico would avoid his least preferred outcome.

Simple as it is, the father-son example illustrates the crucial elements of Calvo's and Kydland and Prescott's analysis of credibility. Often, the interaction between a policymaker and the public is similar to that of Federico and Pablo. Like Pablo, the public makes some decisions whose value depends on subsequent policy actions of the policymaker. Like Federico, the monetary authority may have an incentive to announce policy actions in order to affect the public's decisions and to break its promises once these decisions are made. If the public understands the policymaker's incentives, it will disregard its promises. And this interaction will often result in a bad outcome for society. This is the essence of what Calvo and Kydland and Prescott call the time-inconsistency problem.

Although time inconsistency pervades all aspects of government policy, its application to monetary policy has attracted the most research. A monetary authority, such as the Federal Reserve, typically has as a major objective to deliver low inflation. It may also have other objectives that can be accomplished by creating surprise inflation, that is, inflation rates over and above those previously anticipated by the public. A case in point occurs if one objective is to fight unemployment, as in the studies by Kydland and Prescott (1977) and Barro and Gordon (1983a). These studies assume that firms and workers write contracts before production and sales take place; these contracts stipulate a fixed nominal wage at which workers agree to supply labor at the firms' demand. Then the monetary authority has an incentive to create unexpected inflation that would reduce the real value of wages and induce firms to employ more workers.

If, like Federico, the monetary authority is not bound by its promises, then it will have a credibility problem. The monetary authority would like to promise low inflation but has an ex post incentive to engineer surprise inflation, using whatever policies it has at its disposal, and expand employment.

Can the monetary authority succeed? Arguably, the public is intelligent enough to understand the monetary authority's credibility problem. This premise is, in fact, implied by the more general hypothesis of rational expectations, which was gaining acceptance in macroeconomics when Calvo's and Kydland and Prescott's contributions were published. Rational expectations theory maintains that individuals use efficiently all available information when making decisions. Under the plausible assumption that their information includes knowledge about how monetary policy is chosen, individuals may not believe a promise of low inflation by the monetary authority, just as Pablo discounted Federico's promise in the father-son example. Rather, understanding correctly that the monetary authority will attempt to engineer surprise inflation, individuals will adjust their inflation forecasts upward.

If individuals know that the monetary authority may try to surprise them, what is the outcome, or equilibrium, that will be observed? The answer, first advanced by Barro and Gordon (1983a), is somewhat tricky. The key observation is that a plausible outcome must have the property that the monetary authority does not profit, at the end, from surprising the public. This property must hold because individuals know that the monetary authority will try to create surprise inflation if it can gain from doing so. Therefore, in an equilibrium, expected inflation

must equal actual inflation, and both have to be such that there is no incentive for the monetary authority to create unanticipated inflation.

Since expected and actual inflation must coincide in equilibrium, the monetary authority cannot succeed in its effort to expand employment. Given this restriction, it seems plausible that the monetary authority would choose to keep inflation low. However, time inconsistency means that expectations of low inflation provide an incentive for the monetary authority to create unexpected inflation, which would be incompatible with an equilibrium. Instead, in an equilibrium both expected and actual inflation must be such that that incentive is eliminated. Under plausible conditions the result is inflation that is inefficiently high.[5]

The result is very bad from a social perspective: not only is the monetary authority unable to expand employment, but expectations of high inflation end up being accommodated by monetary policy. In short, monetary policy suffers from an inflation bias because of the time-inconsistency problem.

Like Federico, the monetary authority would be better off if it could somehow bind itself to honor its promises. If that were possible, then the monetary authority would achieve a better outcome by promising to deliver low inflation. Making such a promise would imply giving up on the employment objective, but it would at least succeed in keeping inflation low.

The emergence of a time inconsistency problem when a central bank is concerned with both inflation and employment has been one focus of the literature, and because of its importance the rest of the article will explore this scenario as well. However, the reader should keep in mind that a monetary authority may have to deal with time inconsistency and a resulting inflation bias when it has objectives other than fighting unemployment. For instance, a central bank forced to finance government expenditures through money creation may have an incentive to promise low inflation to maximize the demand for money, which forms the base of the inflation tax, and then to break that promise to increase inflationary revenue. Other examples are not hard to find, suggesting that time inconsistency may be a pervasive feature of monetary policy.

Dealing with Inflation Bias: Delegation and Incentives

It is clear that, in the presence of time inconsistency, a monetary authority would benefit from tying its hands behind its back to enhance its credibility. However, doing so is not so simple. The monetary authority may try to promise or even enact a rule that it will behave "honestly." Such announcements would presumably be no more believable, though, than a promise of low inflation.

What else could be tried? To explore some possibilities, let us return to the father-son example. Obviously, Federico would not suffer from lack of credibility if he did not dislike tattoos. Even with the assumption that Federico hates tattoos, he might obtain desirable results if he were to give up dealing with Pablo directly and delegate Pablo's education to a tutor. In pursuing this solution, Federico should be careful that the tutor's incentives are such that he does not have a credibility problem himself. To ensure the tutor's credibility, Federico has two options. One is to hire a tutor who likes tattoos, on the premise that such a tutor would not be tempted to buy a book rather than paying Pablo for mowing the lawn. The other option is to pay the tutor only according to whether Pablo mows lawns and not according to what he does with the money earned.

Institutional approaches and reputational concerns are both plausible solutions to the time inconsistency problem, and both have weaknesses according to existing theory.

Analogously, a society may try to deal with the time inconsistency of monetary policy by delegating the execution of monetary policy to agents with appropriate incentives. This point was first developed in an important paper by Rogoff (1985). Rogoff studied an idealized economy in which there were well-defined social preferences on inflation-employment combinations. In such an economy, a central banker with the same preferences as those of society would suffer from a credibility problem, as discussed in the previous section. Given this problem, Rogoff's key insight was that this society may not be constrained to choose an individual with the same preferences as itself to conduct monetary policy. Instead, it should choose a person whose distaste for inflation is greater than the social one.

The appointment of a "conservative" central banker mitigates the inflation bias because the public would know that such a person would refrain from using unexpected inflation to expand employment. Accordingly, individuals would reduce their inflation forecasts, and they would turn out to be correct because of the central banker's distaste for inflation. The conservative central banker would not attempt to stimulate employment but

5. *One such condition is that the marginal cost of creating inflation increases the level of inflation while the marginal effect on employment is constant.*

would be able to promise and deliver low inflation. Hence delegating monetary policy to a conservative central banker would improve matters, just as hiring a tutor who likes tattoos would help Federico ensure that Pablo will mow the neighbors' lawn.

Rogoff's analysis may explain why central bankers are often known to be "hawkish" on inflation: according to Rogoff's theory, this position would be a social response to the credibility problem in central banking. More subtly, Rogoff's prescription requires that central bankers be independent of other branches of the government. This independence is needed because the policy choices of a central banker whose preferences are different from those of society must, ex post, be suboptimal from a social perspective. Society therefore would have an incentive to dismiss the conservative central banker when he is about to implement policy, just like Federico would have an incentive to fire the tutor after the lawn is mowed and then buy Pablo a book. This incentive must be held in check for the conservative central banker to be effective.

A society may try to deal with the time inconsistency of monetary policy by delegating the execution of monetary policy to agents with appropriate incentives.

Another option discussed for Federico is that, instead of hiring a tutor who likes tattoos, he may solve his credibility problem by paying the tutor only according to the execution of Pablo's job. Likewise, rather than choosing a conservative central banker, a society may deal with a monetary policy credibility problem by appropriately structuring the rewards and compensation of its central banker, that is, by designing an efficient "contract." What would such a contract look like? Recall that the key implication of time inconsistency is that it induces an inflation bias. Eliminating that bias would seem to require that the central banker be penalized when inflation is high and rewarded when inflation is low. In addition, the contract might stipulate additional rewards or penalties to the central banker depending on other variables such as employment growth.

These questions were first investigated in an influential paper by Walsh (1995a). Walsh obtained several interesting results in the context of the monetary policy model of Barro and Gordon (1983a). He showed that an optimal contract for a central banker would make his compensation depend only on the realized rate of inflation or, alternatively, on the realized rate of money growth. This finding is surprising because one could have conjectured, as in the previous paragraph, that optimal rewards would depend on other variables in addition to inflation or money growth.[6] Also, Walsh showed that the optimal reward structure may resemble an inflation-targeting rule in that the central banker would be rewarded according to how close inflation turned out to be relative to some given values or targets.

Hence Walsh's theory provides a formal justification for inflation targeting and for the recent trend toward assigning central banks the sole objective of maintaining low inflation. That justification is based on the incentives that inflation-based compensation schemes would provide to central bankers. This argument contrasts with others in favor of inflation-based rules for monetary policy, which have emphasized the implications for the distribution of macroeconomic outcomes assuming that the rules will be followed.[7]

As with Rogoff's approach, a necessary condition for Walsh's approach is that the central banker must be independent, in the sense that his contract with society must be respected even if it is beneficial, ex post, to rescind it. In order to deal with time inconsistency, the central banker's contract must induce him not to create unanticipated inflation even if inflationary surprises may be beneficial. If the central banker's contract could be repealed at no cost, the contract would itself become incredible and its effects on the public's expectations would disappear.

Summarizing, the inflation bias caused by time inconsistency may be ameliorated if society can change the incentives of its central banker. This change can be accomplished by choosing a very inflation-averse individual to head the central bank or by designing his contract to discourage him from creating inflationary surprises. The latter approach may resemble a regime of inflation targeting. In both cases, central bank independence emerges as a necessary ingredient to ensure that the change in incentives is effective.

It has been emphasized that incentive-based approaches are feasible provided that society can affect the incentives of its central banker. But doing so may not be possible. If, for example, a government can commit to a particular contract with the head of its central bank, why is it impossible for that government to commit to honor promises of low inflation? The question has no fully satisfactory answer. In the end, the incentives approach depends crucially on the assumption that a society can make some commitments (such as honoring the contract of its central banker) and not others. Such an assumption must ultimately be justified on institutional or political grounds, but on this point theory remains to be developed.[8]

Dealing with Inflation Bias: Rules

Instead of imposing constraints on incentives, a different approach to solving the credibility problem imposes external constraints on the instruments that central bankers can use. The consequences would be

trivial if society could force its central bankers always to honor promises, but research along these lines assumes that imposing such a stringent constraint is not feasible. Instead, it is assumed to be feasible to impose other, less than perfect rules. Then the interesting question is to investigate the implications of those rules for equilibrium outcomes.

Consider again the father-son problem, assuming that Federico deals directly with Pablo. It may be impossible for Federico to credibly promise that he will give Pablo money to pay for a tattoo as his reward for mowing the neighbors' lawn. Nevertheless, Federico may instruct the neighbors to pay Pablo directly; this approach would prevent Federico from using the money to buy books instead. This arrangement may be a good idea in spite of, or precisely because, everyone knows that Pablo will get a tattoo if he gets the money.

Analogously, a society may be able to restrict the actions of its monetary authorities so as to alleviate the inflation bias caused by time inconsistency. The commitment to fixed exchange rates in European countries has been justified in this fashion (see Giavazzi and Pagano 1988). If a country is committed to a fixed exchange rate, it becomes difficult for the central bank to engineer inflation surprises, as they are likely to put downward pressure on the country's currency. This constraint tends to reduce the inflation bias since the public understands its consequences for monetary policy.

This kind of reasoning also provides a justification for simple monetary rules, such as a constant money growth rule. These procedures are interpreted as constraints on the choices available to central bankers, designed with the purpose of ameliorating inflation bias by preventing inflation surprises.

The conclusion is that, provided society can commit to at least some rules, the imposition of rules may help deal with credibility. This view may help justify some rules, such as fixed exchange rates, that would otherwise be irrelevant or even counterproductive.

As with incentive-based approaches, a key question is what policy choices can and cannot be ruled out. In justifying a fixed exchange rate regime, the implicit assumption is that the monetary authority can commit to fixing exchange rates but not to honoring promises of low inflation. Why is there a difference? One answer is that, because of institutional reasons, some commitments are harder to break than others. This argument carries some force for fixed exchange rate regimes, which may require international agreements that are costly to ignore. However, even in the case of fixed exchange rates the argument is grounded on an institutional factor and not completely satisfying.

An alternative answer holds that society can in fact commit to rules, but only imperfect ones. This limitation exists because there is incomplete knowledge about the nature of shocks that may hit the economy. In this view, espoused most prominently by Flood and Garber (1989), the assumption that society can commit to some rules but not to honoring its promises approximates the fact that no rule can be written that takes into account every possible kind of disturbance to the economy. Is this argument convincing? That all rules are imperfect is not controversial. However, Flood and Garber's interpretation amounts to assuming that policymakers cannot use standard probability theory to describe the likelihood of some relevant macroeconomic shocks. This assumption is problematic, for it makes it very hard to solve the model in a convincing way. In particular, how are individuals supposed to make investment and portfolio decisions in such an environment? Flood and Garber assume that agents' choices are based on rational expectations, that is, on full knowledge of the structure of the economy. But such knowledge must include the probabilities of all the shocks, and therefore its availability is inconsistent with Flood and Garber's interpretation of the limitations of policy rules.

The monetary authority may eliminate the inflation bias by developing a reputation for honoring its announcements of low inflation.

The discussion in this article implies that there is no satisfactory justification for assuming that a society can commit to some rules but not to others. Such an assumption makes the theory interesting but may also be its main weakness.

The Role of Reputation

So far the emphasis has been on institutional responses to the problem of policy credibility. Since the creation and enforcement of appropriate institutions may be difficult and costly, one should ask whether

6. As Walsh noted, this result depends on the assumption that the central banker cares about not only his compensation but also social welfare. Also, the result hinges on a particular property of the Barro-Gordon model—that the magnitude of the inflation bias is not affected by macroeconomic shocks. Whether the result holds under more general assumptions is the subject of current research.

7. For an example of this kind of argument, see Taylor (1993).

8. And, indeed, this point has been identified as a major weakness of the incentives approach (see McCallum 1995).

there are other ways to deal with the credibility problem. One alternative exists provided that the monetary authorities and the public interact for sufficiently long periods. In such cases, the monetary authority may eliminate the inflation bias by developing a reputation for honoring its announcements of low inflation, as discussed briefly above.

In the father-son example, it is unlikely that Federico would be left unpunished if he breaks a promise to Pablo. In real life Federico and Pablo would have to face each other for many years. Federico may therefore be deterred from reneging on his promises by the fear that Pablo will not believe subsequent promises. Somewhat paradoxically, Federico's fear of losing his reputation vis-à-vis Pablo is in fact useful. Federico may be able to credibly promise Pablo a tattoo in exchange for mowing the neighbors' lawn if Pablo believes that Federico wants strongly enough to be able to make credible promises in the future.

Central bank independence plays a role in dealing with time inconsistency but only as a complement to more fundamental arrangements intended to bind central bankers to honor their promises.

The same considerations apply to monetary policy. It is likely that a monetary authority that today makes and breaks a promise of low inflation will be unable to credibly promise low inflation in the future. This incentive may be powerful enough to deter the authority from reneging on its current promise because, as the discussion has shown, the ability of making credible promises is socially valuable.

These ideas were first discussed in the context of monetary policy by Barro and Gordon (1983b). They analyzed a simple version of the monetary model in Kydland and Prescott (1977), the main difference being the assumption that the monetary authority and the public interacted for many periods. One of the outcomes of that interaction, Barro and Gordon found, was that the monetary authority acted as if it were able to make binding promises in every period, provided it were patient enough.[9]

In spite of its importance, further study of the role of reputation was hindered for several years by the technical issues involved in analyzing the long-term relationship between a central bank and the public. The main problem is that such analysis quickly leads to a problem of infinite regress. Describing an outcome of a long-term interaction requires specifying not only what happens if the central bank breaks a promise of low inflation but also what happens if it breaks another promise after the first one and then a third promise, and so forth. In fact, even the very concept of equilibrium, that is, of the plausible outcomes of a model, is not obvious.

Very recently, however, new methods have appeared that promise a drastic reconsideration of models of reputation. Chari and Kehoe (1990) and Stokey (1991) provide a convincing definition of equilibria in macroeconomic models of long-term relationships. In addition, these two papers present a general method for identifying the complete set of equilibrium outcomes of many such models. Although that method turns out to be difficult to apply, recent studies by Chang (1998) and Phelan and Stachetti (1997) have shown how the solution of such models can be drastically simplified, thus greatly broadening the scope of the theory of reputation.

Chang and Phelan-Stachetti exploit the fact that, in some sense, tomorrow will be very similar to today. To see this concept, recall that an outcome of the long-term relationship between the central bank and the public involves a description of what will happen if the central bank breaks a promise today, tomorrow, or the day after tomorrow, and thus ad infinitum. This kind of analysis can be exceedingly complex. However, Chang and Phelan-Stachetti show that an equivalent description can be obtained by focusing on the central bank's decision problem today, after any possible history of (possibly broken) promises, with the understanding that tomorrow's problem will be just like today's (except that the relevant history will be a little different). This approach effectively reduces the analysis to a two-period problem, involving only today and tomorrow, and hence eliminates the infinite regress problem.

The papers by Chang (1998) and Phelan and Stachetti (1997) discuss in detail the theoretical advantages of their formulation. Interestingly, they find that in any equilibrium monetary policy must follow "rules," even in the absence of external mechanisms to enforce them. The intuition is as follows. A crucial part of Chang's and Phelan-Stachetti's method is that, at any point in time, the whole history of the economy can be summarized by a small number of "state" variables. An implication is that any equilibrium has the property that monetary policy and market outcomes depend only on those variables and not on calendar time. Since monetary policy is, in any equilibrium, optimally chosen by the monetary authority, this reasoning reveals that monetary policy is governed by a relatively simple relationship between the state variables and policy instruments, that is, by a rule. Although the nature and properties of the resulting rules remain to be investigated, this finding is important because it means that observing that monetary policy follows rules should be the norm and not the exception.

In addition, Chang's and Phelan-Stachetti's studies imply that models of reputation in monetary policy can be analyzed by computational methods that many others

thought were inapplicable. Chang's study, in particular, analyzes the model of Calvo's (1978) original contribution and shows that reputational considerations may imply that one of the model's outcomes is indistinguishable from the best the central bank can do when it can commit perfectly to its promises. That is, the time-inconsistency problem may not prevent the attainment of socially optimal outcomes.

While the results just described are suggestive, it is too early to conclude that the long-term nature of the interaction between a central bank and the public implies a solution to the credibility problem. Other, more realistic models of monetary policy need to be investigated. More importantly, that the central bank can implement its most preferred policy is only one of many possible outcomes. In most models it remains possible that reputational considerations will not be enough to convince the public that a central bank will, in fact, honor its promises. Consider, for example, what would happen if the central bank were to assume that reputation will never matter for the public's behavior. In such a case, the central bank might behave myopically and, in general, try to cheat on the public at all times. This behavior may in turn validate the public's belief that the central bank will not attempt to develop a "good" reputation. The outcome would be that the role of reputation would not solve the time-inconsistency problem, even if the central bank and the public face each other indefinitely.

Since the theory of reputation implies that the central bank interaction with the public may have multiple outcomes, determining which outcome will occur becomes a key issue. Unfortunately, existing studies do not provide a satisfactory answer, and at this point the presumption that reputational effects eliminate the inflation bias caused by time inconsistency is based on optimism rather than theory.

Some Comments on the Empirical Evidence

A natural reaction to the theoretical discussion above is to turn to empirical evidence to check whether the credibility problem of central banks is, in fact, a problem. Unfortunately, testing the various theories described in this article has proved to be very difficult.

One of the sources of difficulty is that it is impossible to measure credibility directly. To see this problem, consider testing the key proposition that more credible central banks deliver lower inflation. What dimension identifies a central bank as "more credible"? In the theoretical discussion, credibility is a central bank's ability to make binding promises. How can such ability be observed, let alone measured, in the real world?

Because of these difficulties, existing empirical studies have by and large focused on testing a different but related proposition: that more independent central banks deliver lower inflation. This approach can be seen as an indirect test of the theory since, as the discussion noted earlier, central bank independence may emerge as part of society's attempt to eliminate the inflation bias caused by lack of policy credibility. If such an attempt is successful, one should expect a high degree of central bank independence to be associated with low inflation.

It is likely that a monetary authority that today makes and breaks a promise of low inflation will be unable to credibly promise low inflation in the future.

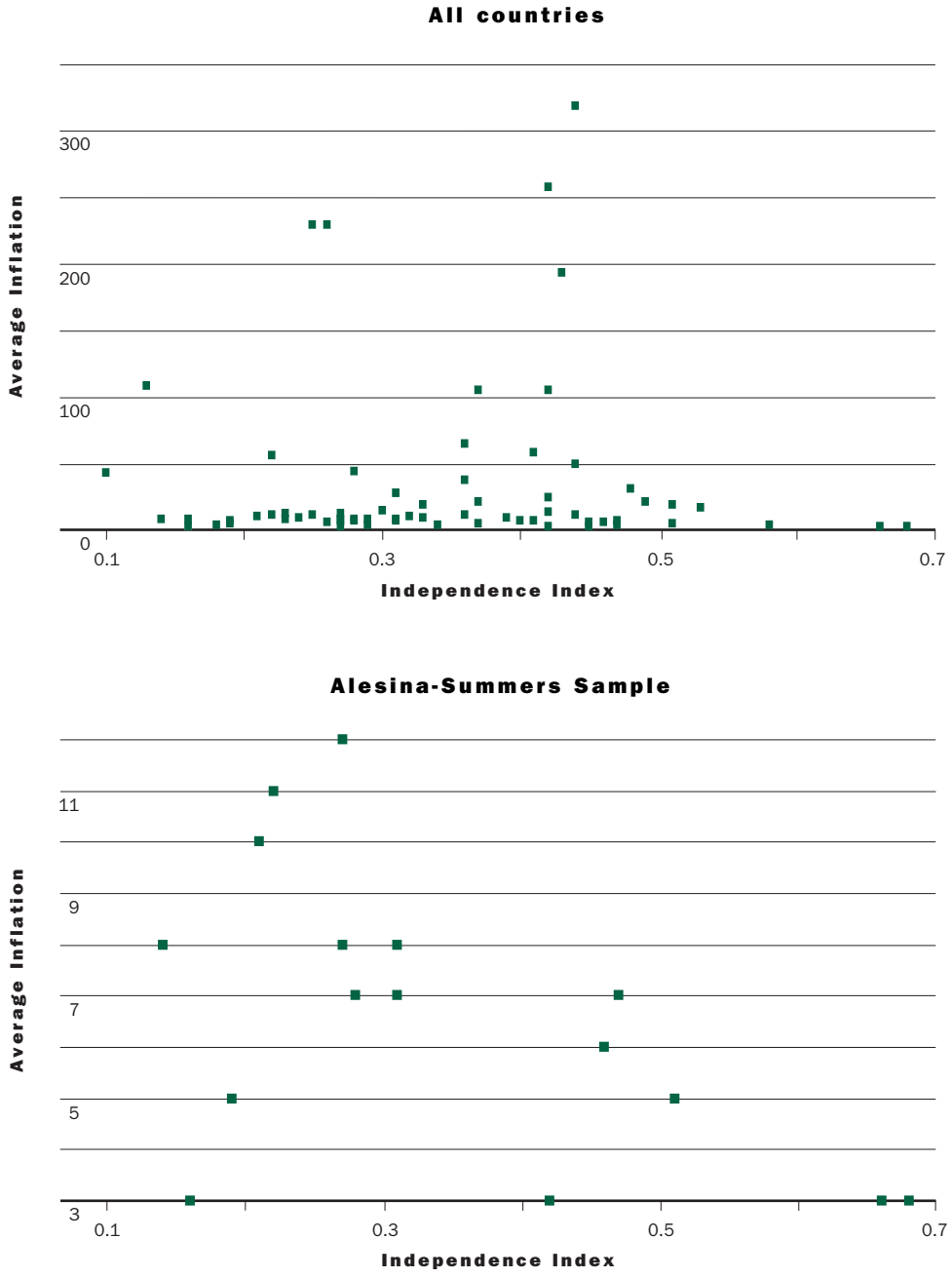
The change of focus from central bank credibility to central bank independence is useful because independence is typically expressed in many indicators found in legal documents and central bank statutes. In some cases, there is little disagreement on when such indicators signal more or less independence; for instance, most people would agree that a central bank whose chairman can be fired at the will of the president of the country is less independent than one whose chairman cannot be so easily dismissed.

The most comprehensive attempt to quantify central bank independence is given by Cukierman (1992). He rates the central banks of several countries in different decades according to four measurable dimensions of central bank independence. The first concerns the procedures governing the appointment, tenure, or dismissal of central bankers. For example, Cukierman rates a central bank whose head is appointed by the executive branch of the government as less independent than one whose chairperson is appointed by the legislature, which in turn is rated less independent than one whose head is chosen by the central bank's board. Likewise, he considers a central bank to be more independent the longer its chairperson's statutory tenure is.

The second dimension is related to the formulation of monetary and fiscal policy. Cukierman gives a high independence rating to central banks that can decide on monetary policy without interference from the executive or legislative branches. In contrast, he gives lower ratings to central banks that must obey their government's decisions about the formulation and execution of monetary policy.

9. *The monetary authority's degree of impatience is important in Barro and Gordon's analysis because the punishment they considered for a government that reneges on today's promise is the loss of reputation in the future; this punishment carries less force if the future is discounted more heavily.*

CHART 1 Central Bank Independence and Inflation, 1980–89



A third dimension of independence has to do with the goals that a central bank is instructed to pursue. Central banks whose sole objective is to ensure low inflation are given high independence ratings. If the central bank's mandate includes other objectives, such as pursuing full employment, that bank is given a lower rating. It can be argued that whether price stability is the central bank's only objective has little to do with the usual meaning of independence. Cukierman's rationale is that the preeminence of price stability among a central bank's possible objectives measures society's willingness to have a conservative central banker.

The fourth and final dimension of independence lies in the extent to which a central bank is required to finance government deficits. The easier the terms are under which a central bank is required by law to finance government deficits, the lower its independence rating is.

Since each available indicator is likely to convey some information about central bank independence, it is useful to include all of them in empirical work. Cukierman (1992) and others do so by constructing indices of central bank independence. Each index is essentially a weighted average of many indicators. Since the weights can vary from study to study, the construction of an independence index involves some subjectivity. However, the conclusions obtained in the existing literature do not seem to depend on the use of a particular index. Those conclusions, therefore, merit attention.

Recent research has underscored the difficulty of obtaining a tight relationship between measures of central bank independence and inflation, as the top panel of Chart 1 illustrates. Each point in the chart represents a country's central bank independence, measured along the horizontal axis, against its long-term inflation, measured along the vertical axis. Cukierman's index is a proxy for central bank independence, and the annual percentage change in the consumer price index is used for long-term inflation; both variables refer to the 1980–89 decade.

A glance at the top panel of Chart 1 suggests the absence of a systematic link between central bank independence and inflation. This conjecture is confirmed by formal statistical tests, which reveal that increases in the Cukierman index are associated with mild increases in inflation, although the association is not significant. Ordinary least squares applied to the data in the top panel of Chart 1 yields the following estimated equation: $INF = 30.27 + 15.05\ CBI$, where *INF* denotes 1980–89 average inflation and *CBI* denotes Cukierman's index for each country. The *t*-statistic associated with the *CBI* coefficient is 0.23, which is quite consistent with the hypothesis that the *CBI* coefficient is zero. Since an increase of Cukierman's index expresses a higher degree of central bank independence, the data in the top panel of Chart 1 suggest that the empirical relationship between independence and inflation is the opposite of that predicted by the theory of time inconsistency.

The above finding seems to contradict the hypothesis that central bank independence translates into lower inflation. Belief in that hypothesis has become widespread after the publication of news stories discussing studies that seem to confirm it.[10] The difference between those studies and the results reported here can be explained easily. For a small subset of developed countries, greater central bank independence seems to be associated with lower inflation, as the theory predicts. To illustrate the point, the bottom panel of Chart 1 plots the same data as in the top panel but for only a subset of developed countries. In fact, the countries included in the bottom panel are the ones studied in an influential paper by Alesina and Summers (1993). For this subset of countries, the bottom panel of Chart 1 suggests the existence of a negative relation between Cukierman's index and inflation, a conjecture that is confirmed by formal statistical tests. For the sample of the bottom panel, ordinary least squares yields the following equation: $INF = 9.79 - 9.11\ CBI$. The *t*-statistic associated with the *CBI* coefficient is –2.36, which is inconsistent with the hypothesis of a zero *CBI* coefficient at conventional significance levels.[11]

At this point the presumption that reputational effects eliminate the inflation bias caused by time inconsistency is based on optimism rather than theory.

The conclusion is that greater central bank independence seems to have no beneficial impact on inflation, except perhaps for a small group of developed countries. How should these findings be interpreted? A possible reaction is to keep believing that independence is conducive to lower inflation, blaming shortcomings in empirical procedures for failing to confirm that belief. It has been argued—for example, by Cukierman (1992)—that the problem is one of poor measurement. The independence indices used in Chart 1, as well as in most of the literature, capture only the legal aspects of central bank independence. The real degree of independence may depend on other, nonlegal variables that are hard to quantify. The solution seems to lie in finding alternative, more accurate measures of central bank independence; research on that front is still under way.[12]

10. Examples are "Role Shifts for Central Bankers," New York Times, *November 15, 1994, sec. D, and "Divorcing Central Banks and Politics: Independence Helps in Inflation Fight,"* New York Times, *May 7, 1997, sec. D.*

11. This finding need not imply that central bank independence helps lower inflation. An alternative explanation is that countries that have stronger anti-inflationary postures tend to be more conservative with their central bank arrangements. This "reverse causality" view is proposed by Posen (1995).

12. Cukierman (1992) observes, for instance, that in some countries the average tenure of central bank presidents is much shorter than the legal tenure period, which is one of the variables summarized in independence indices. Accordingly, Cukierman argues that central bank independence can be measured more accurately in less developed countries by the turnover ratio of their central bank heads.

An alternative reaction to the empirical findings summarized by Chart 1 is that the theory implies that central bank independence is associated with lower inflation only under narrow conditions that may not hold in practice. As was discussed earlier, the inflation bias problem associated with time inconsistency may be solved if central bankers develop a good reputation with the public. If reputation does in fact work, one should not expect to find any systematic relationship between central bank independence and inflation, and hence the empirical facts reported earlier are not a puzzle.

A more pessimistic view is that central bank independence is only a necessary but not a sufficient condition for eliminating the inflation bias caused by time inconsistency. According to the theory discussed earlier in this article, central bank independence plays a role in dealing with time inconsistency but only as a complement to more fundamental arrangements intended to bind central bankers to honor their promises. It may be the case that central bank independence emerges for reasons not related to those institutions, but in and of itself it does not help solve the time-inconsistency problem and, therefore, does not result in lower inflation.

Conclusion

This article has reviewed the problem of time inconsistency of monetary policy and its possible solutions. The theory of time inconsistency emphasizes that, if a central bank cannot credibly commit to honor announcements of low inflation, expected and actual inflation will be larger than if such a commitment could be made. In other words, time inconsistency leads to an inflation bias.

The discussion considers how some currently fashionable institutions such as central bank independence and price stability rules may emerge as attempts to minimize the inflationary consequences of time inconsistency. But it also argues that there may be no need for such institutions. The empirical evidence reviewed here did not provide strong confirmation of the hypothesis that central bank independence lowers inflation. This empirical failure may reflect that the time inconsistency bias has been solved by reputational considerations, as suggested by recent theoretical advances. Alternatively, it may be the case that the degree of central bank independence is determined by reasons other than eliminating inflation bias.

REFERENCES

Alesina, Alberto, and Lawrence H. Summers. 1993. "Central Bank Independence and Macroeconomic Performance." *Journal of Money, Credit, and Banking* 25 (May): 151–62.

Barro, Robert J., and David B. Gordon. 1983a. "A Positive Theory of Monetary Policy in a Natural Rate Model." *Journal of Political Economy* 91:589–610.

———. 1983b. "Rules, Discretion, and Reputation in a Model of Monetary Policy." *Journal of Monetary Economics* 12:101–21.

Calvo, Guillermo A. 1978. "On the Time Consistency of Optimal Policy in a Monetary Economy." *Econometrica* 46 (November): 1411–28.

Chang, Roberto J. 1998. "Credible Monetary Policy in an Infinite Horizon Model: Recursive Approaches." *Journal of Economic Theory,* forthcoming.

Chari, V.V., and Patrick Kehoe. 1990. "Sustainable Plans." *Journal of Political Economy* 98:783–802.

Cukierman, Alex. 1992. *Central Bank Strategies, Credibility, and Independence.* Cambridge, Mass.: MIT Press.

Flood, Robert P., and Peter Garber. 1989. "Monetary Policy Strategies." *IMF Staff Papers* 36:612–32.

Giavazzi, Francesco, and Marco Pagano. 1988. "The Advantage of Tying One's Hands: EMS Discipline and Central Bank's Credibility." *European Economic Review* 32 (June): 1055–82.

Kydland, Finn E., and Edward C. Prescott. 1977. "Rules Rather than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy* 85 (June): 473–92.

McCallum, Bennett T. 1995. "Two Fallacies Concerning Central Bank Independence." *American Economic Review* 85 (May): 207–11.

Phelan, Christopher, and Ennio Stachetti. 1997. "Subgame Perfect Equilibria in a Ramsey Taxes Model." Unpublished manuscript, Northwestern University.

Posen, Adam. 1995. "Declarations Are Not Enough: Financial Sector Sources of Central Bank Independence." In *NBER Macroeconomics Annual 1995*, edited by Ben Bernanke and Julio Rotemberg. Cambridge, Mass.: MIT Press.

Rogoff, Kenneth. 1985. "The Optimal Degree of Commitment to an Intermediate Monetary Target." *Quarterly Journal of Economics* 100 (November): 1169–89.

Stokey, Nancy. 1991. "Credible Public Policy." *Journal of Economic Dynamics and Control* 15:627–56.

Taylor, John B. 1993. "Discretion versus Policy Rules in Practice." *Carnegie Rochester Conference Series in Public Policy* 39 (December): 195–214.

Walsh, Carl E. 1995a. "Optimal Contracts for Central Bankers." *American Economic Review* 85 (March): 150–67.

———. 1995b. "Is New Zealand's Reserve Act of 1989 an Optimal Central Bank Contract?" *Journal of Money, Credit, and Banking* 27 (November): 1179–91.

A Dynamic Multivariate Model for Use in Formulating Policy

TAO ZHA
The author is an economist in the macropolicy section of the Atlanta Fed's research department. *He is indebted to Eric Leeper for instrumental suggestions. He is also grateful to Rob Bliss, Jerry Dwyer, Bob Eisenbeis, Frank King, Will Roberds, Mary Rosenbaum, Ellis Tallman, and Dan Waggoner, whose detailed comments have significantly improved the article's argument and exposition. Bryan Acree and Jeff Johnson provided able research assistance.*

ON MARCH 25, 1997, THE FEDERAL OPEN MARKET COMMITTEE (FOMC) RAISED ITS KEY SHORT-TERM INTEREST RATE TARGET—THE FEDERAL FUNDS RATE—BY 25 BASIS POINTS. THE *WALL STREET JOURNAL* CALLED THE MOVE CHAIRMAN ALAN GREENSPAN'S "PREEMPTIVE STRIKE AGAINST INFLATION" (WESSEL 1997). ACCORDING TO GREENSPAN, THE FOMC "BELIEVES IT IS CRUCIAL TO KEEP INFLATION CONTAINED IN THE NEAR TERM AND ULTIMATELY TO MOVE TOWARD PRICE STABILITY" (1997A, 1). THE FOMC DESCRIBED THIS INCREASE "AS A PRUDENT STEP THAT AFFORDS GREATER ASSURANCE OF PROLONGING THE CURRENT ECONOMIC EXPANSION BY SUSTAINING THE EXISTING LOW INFLATION ENVIRONMENT THROUGH THE REST OF THIS YEAR AND NEXT" (WESSEL 1997).

The notion of "preemptive strike" or "prudent step" connotes the most important part of policy making: the process of looking forward. Because the Federal Reserve's monetary policy has effects on the overall economy only through long and variable delays, policymakers must look forward to forecast, to the best of their abilities, how today's policy actions will affect economic conditions such as inflation in the future. This process of anticipating the future is indispensable in formulating sound monetary policy (see, for example, Cecchetti 1995, King 1997, and Blinder 1997).

The Humprey-Hawkins Act has set out multiple objectives for the Federal Reserve, including balanced growth and stable prices (Board of Governors 1994). A policy action by the Fed consists of using any one of various instruments, such as the federal funds rate and different measures of money, to pursue its multiple objectives. However, to provide clearer analysis this article characterizes monetary policy actions more narrowly as changes in the federal funds rate and the discussion concentrates on only one of the Federal Reserve's objectives—keeping inflation, as measured by the consumer

price index (CPI), low and stable. In such a framework, one aspect of effectively advising policymakers is to provide a forecast of how inflation outlook changes if the Federal Reserve adopts different paths of the federal funds rate over the next two or three years. By consulting a menu of such projected outcomes, called policy projections, policymakers can decide which particular policy actions are most likely to keep inflation around the level commensurate with their objective.

Policy projections are essential in helping policymakers decide on policy actions. Unfortunately, obtaining an accurate estimation of such projections is a daunting task. Because the projections are based on various forecasts under different scenarios—here, alternative federal funds rate paths—the first and critical step is to develop good forecasting models (Sims 1980). It is therefore the purpose of this article to present a forecasting model that seems to overcome conceptual and empirical difficulties encountered in other models and promises to provide policymakers with a more useful tool for anticipating effects of policy.

The model, one of a class of models called dynamic multivariate models, introduces new techniques that offer two distinctive advantages. One is the ability to forecast the values of key macroeconomic variables such as inflation and output beyond a period over which these values are known, on the assumption that the trends followed within the period continue beyond it. These extrapolated forecasts are known in technical jargon as out-of-sample forecasts. The model's other advantage is its explicit structure that allows empirically coherent ways to assess the uncertainty of forecasts through error bands. These error bands are constructed so that there is a two-thirds probability that actual outcome is contained within the band.

The article first discusses dynamic multivariate modeling in general and reviews other approaches to forecasting. The discussion then turns to the model itself. After describing the specifics of the model, the article presents the model's point forecasts through the 1980s and 1990s. These forecasts represent the scenarios most likely to develop. Finally, the article shows how to use probability distributions to gauge forecast errors.

Dynamic Multivariate Modeling

The term *dynamic* means that economic variables influence one another through variable lags over time. For example, today's change in the federal funds rate will have consequences on the path of inflation in a year or two. The term *multivariate* implies that a set of multiple variables are examined together, not separately, in one framework. By *dynamic multivariate models* this article means a class of models that are designed to capture, in a single framework, joint movements and dynamic patterns in an array of multiple key macroeconomic variables over a particular period of time. (Technical details are discussed in Box 1 in relation to the specific model presented here.)

Other Approaches. Before explaining the key aspects of dynamic multivariate modeling, it is perhaps useful to review briefly two other approaches to forecasting and policy analysis. One approach is to use rules of thumb. Rules of thumb are often used in actual policy discussions because they may be based on theoretical work and thus can provide compelling stories to policymakers. Unfortunately, they are generally insufficient for characterizing the actual economy, and therefore forecasts derived from these rules are likely to be quantitatively unreliable. For example, one rule of thumb often referred to in the popular press is the Phillips curve relationship, which implies that whenever the unemployment rate is low (high), inflation will soon rise (fall).[1] Chart 1 displays annual inflation and the annual unemployment rate from 1960 to 1996. As the chart shows, there were times when inflation and unemployment tended to move in the same direction, not in opposite directions. For instance, from the early to mid-1970s, rising unemployment was coupled with rising inflation; from 1982 to 1986 both inflation and the unemployment rate fell. During other times inflation and unemployment moved in quite different fashions. Consider 1992–96, for example. During this period, the unemployment rate fell steadily but inflation stayed almost flat. If one used the negative relationship between inflation and unemployment in the 1987–91 period to predict inflation, the result would be to overpredict inflation for 1992–96.[2]

The dynamic multivariate model presented in this article provides a useful tool for gauging future uncertainty and an empirically consistent way to update forecasts.

Another example of rules of thumb is the bivariate relationship between inflation and the growth rate of money. A number of economists (for example, Friedman 1992) have argued that the M2 growth rate in particular appears to have a stable relationship to inflation. Chart 2 displays time-series patterns of inflation and the M2

1. *A.W. Phillips first noted such a relationship in 1958. His original study examined a temporary trade-off between changes in nominal wages and the unemployment rate in the United Kingdom over a period from 1861 to 1957.*
2. *The literature presents several versions of the bivariate relationship between unemployment and inflation. For critical discussions consult, for example, Chang (1997), Espinosa and Russell (1997), and Staiger, Stock, and Watson (1997).*

Details of the Model

This box, heavily drawn from Sims and Zha (1998), describes the important features of the model that is used to produce the results presented in Charts 6–10. The dynamic multivariate model takes the following simultaneous equations form:

$$y(t)A(L) = \varepsilon(t), t = 1,...,T, \quad (1)$$

where $A(L)$ is an $m \times m$ matrix polynomial of parameters in lag operator L, $y(t)$ is a $1 \times m$ vector of observations of m variables at time t, and $\varepsilon(t)$ is a $1 \times m$ vector of independently, identically distributed (i.i.d.) structural shocks so that

$$E\varepsilon(t) = 0, E\varepsilon(t)'\varepsilon(t) = \underset{m \times m}{I}. \quad (2)$$

Note that T is the sample size. To estimate system (1), the likelihood function is multiplied by a probability density function. This probability density, formally a Bayesian prior distribution, aims at eliminating the undesirable problems associated with the estimation. These problems are discussed in detail below.

The number of parameters in $A(L)$ grows with the square of the number of variables in system (1). Given the short period of macroeconomic data after World War II, traditional, ordinary least squares (OLS) estimation of a large model (for example, the eighteen-variable model studied by Leeper, Sims, and Zha 1996) becomes imprecise because of relatively low degrees of freedom and a large number of parameters. Thus, models used in macroeconomics are often of small size (say, six variables). For small models like the six-variable model presented in this article, error bands on the OLS estimates of parameters are usually tight, and thus quantitative analysis from these models can be informative. Nonetheless, when a model is used for out-of-sample forecasting, one can no longer take comfort in "good" in-sample properties of the OLS estimates. Three major problems prevent reasonable out-of-sample forecasting, especially over long horizons (such as two or three years out).

The first problem is a familiar one: overfitting. Because of a large number of parameters, the model tends to fit the sample unrealistically well but fails badly for out-of-sample forecasting.[1] To see how unbelievable the overfitting problem could become, Chart A displays actual values and in-sample (*not* out-of-sample) forecasts of the stock of M1 from January 1960 to March 1996. These in-sample forecasts, drawn directly from Sims and Zha (1998), are made as of 1959:12 from the estimated model (using the data from 1959:7 to 1996:3) without any prior distribution (that is, with OLS estimates). As shown in Chart A, one could, in 1959, predict with almost perfect precision the level of M1 stock in 1996—an incredible outcome.

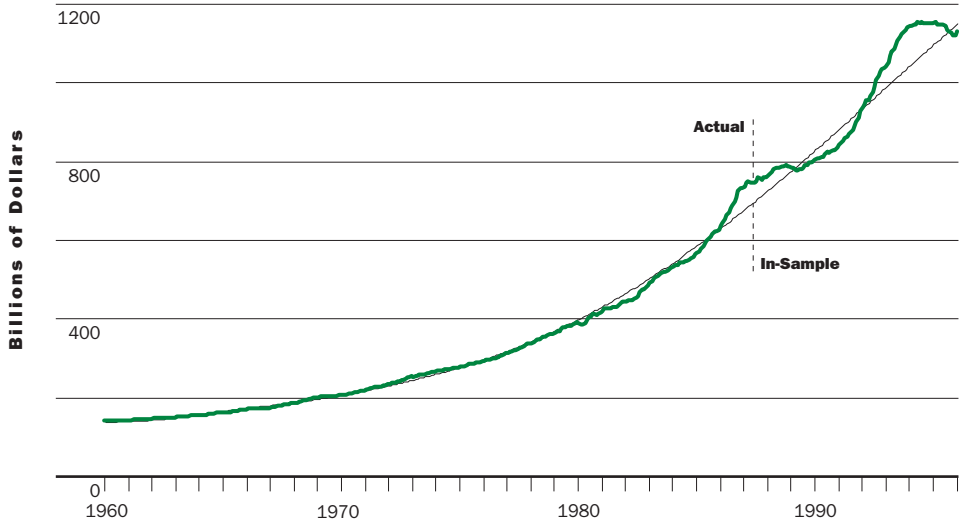
Another aspect of overfitting, which has not been addressed in the textbooks, is an unreasonable extraction of business cycles into deterministic components (see Sims and Zha 1998 for technical details). This undesirable feature may have contributed to findings about substantial differences in OLS estimates across different subsamples. It may distort long-run relationships among variables in the model as well. To deal with these overfitting problems, the model used here, following Sims and Zha (1998), uses priors that favor unit roots and cointegration.[2] At the same time, the model avoids imposing exact, but likely spurious, unit roots and cointegrated relationships with a probability of one.

The third problem relates to low degrees of freedom in most macroeconomic models. Typically, OLS estimates tend to produce large coefficients on distant lags and erratic sampling errors. One of the prior distributions used in the model here is to downweight the influence of distant lags or the unreasonable degree of explosiveness. This prior distribution is essential for ensuring reasonable small-sample properties of the model, especially when degrees of freedom are relatively low.

The prior distributions used here do not intend to encompass all briefs that are likely to improve out-of-sample forecasts. Rather, they reflect some widely held briefs that are likely to be uncontroversial. In this sense, the prior distributions are of a reference nature, and such an approach closely follows the likelihood principle.

1. *Dynamic multivariate models are not the only types that produce overfitting. This problem is common across many empirical models (see Diebold 1998b).*
2. *From a different perspective, Christofferson and Diebold (1997) discuss why cointegrated relationships are important for short-term forecasts.*

CHART A
Actual and Forecast M1 Monthly Series
(1960:1–1996:3)



Source: Sims and Zha 1996.

growth rate from 1960 to 1996. The M2 growth rate reached a peak three times—in 1972, 1976, and 1983. But the path of inflation after each peak was quite different. Clearly, past M2 growth rates predict future inflation through variable lags, and there are no regular patterns.

Another approach to forecasting is to link forecasts of macroeconomic variables to a large array of other variables through econometric techniques. This approach usually involves many strong assumptions or judgmental adjustments. Large-scale structural econometric models are examples of this approach. The goal of these models is to not only provide forecasts of key macroeconomic variables but also examine in detail many different sectors of the economy (Diebold 1998a). Because of their detailed, intricate nature, however, these models are often difficult to produce and evaluate independently. Furthermore, strong assumptions contained in these kinds of models, such as the Phillips curve relationship, may be at odds with the data. Judgmental adjustments consequently play roles in the model's outcomes from period to period. Such periodical adjustments make it difficult to gauge the quality of the model itself.

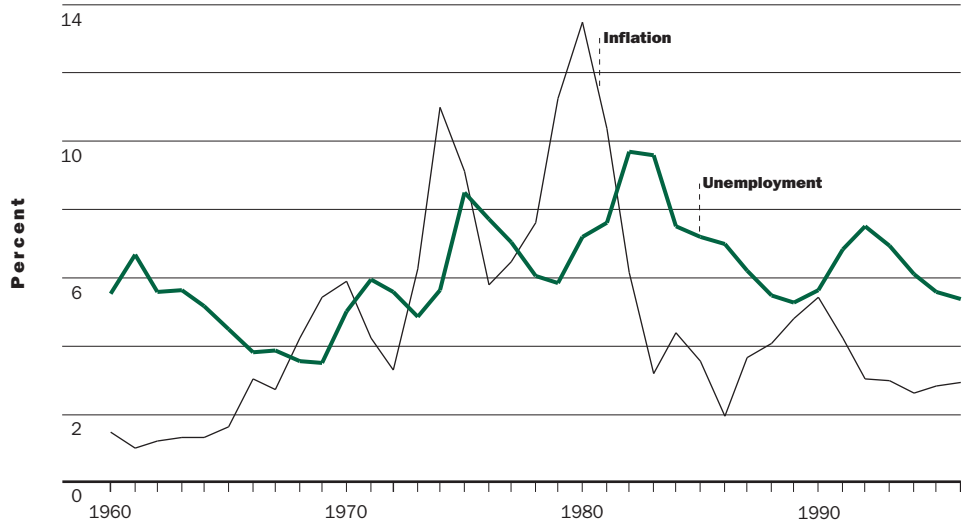
Distinctive Aspects of Dynamic Multivariate Modeling. Dynamic multivariate modeling offers a different approach. It is not designed to study every detail of the economy. Rather, it is designed to capture only essential elements so that the model can be readily understood and reproduced. It is closely connected to modern economic theory and usually involves only six to eight variables.[3] After the model—the array of variables, the lag length, and other assumptions—is set up, forecasts from the model will not be altered from period to period on the basis of judgments or assumptions outside the model itself. Thus, the model can be evaluated objectively.

At the same time, dynamic multivariate modeling has complex structures in the sense that it allows both contemporaneous and dynamic interactions among the macroeconomic variables. In relation to rules of thumb, dynamic multivariate models capture the relationships implied by these rules if such relationships exist in the data. In contrast to large-scale models, dynamic multivariate modeling avoids imposing strong assumptions that may be at odds with the data. Consequently, both the Federal Reserve's complex behavior and the public's expectations about future policy actions are implicitly embedded in dynamic multivariate models.

More important, dynamic multivariate modeling provides empirically coherent ways to assess the uncertainty about forecasts (Sims and Zha 1998). All forecasts have errors. The errors usually come from two sources—uncertainty about model parameters and uncertainty emanating from exogenous shocks (that is, those that cannot be predicted by the model). Dynamic multivariate

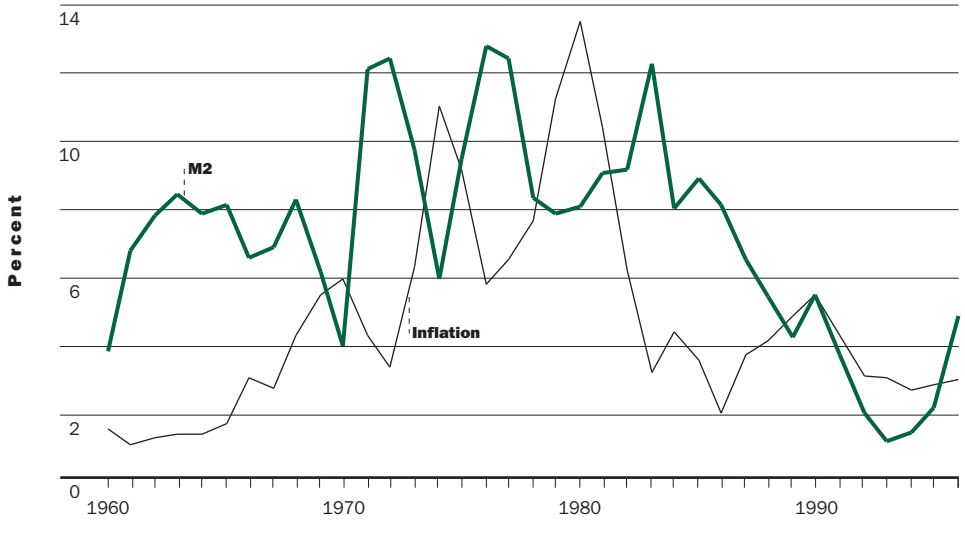
3. See, for example, Diebold (1998a) and Sims and Zha (1996) for detailed discussions.

CHART 1 Annual Inflation and Unemployment Rates, 1960–96



Source: See Box 2.

CHART 2 Annual Inflation and M2 Growth Rates, 1960–96



Source: See Box 2.

modeling lays out a probabilistic structure that takes both types of uncertainties into account explicitly. When probability distributions or error bands are attached to point forecasts, policymakers will be well informed of the likelihood of future inflation.

The Model

The dynamic multivariate model used in this article employs monthly data with the six key macroeconomic variables often used in the literature: the federal funds rate, the stock of M2, the consumer price index, real (inflation-adjusted) gross domestic product, the unemployment rate, and an index of commodity prices (see Box 2 for a precise description of the data set). The data begin at 1959:1 and end at the time when the forecast is made. The model allows these variables to interact with one another both simultaneously and through lags.[4] The lag length is thirteen months, meaning that variables in the past thirteen months are allowed to affect those in the current month.

BOX 2

Data Description

The model uses monthly data from 1959:1 to 1997:9 for six macroeconomic variables:

CPI. Consumer price index for urban consumers (CPI-U), seasonally adjusted. Source: Bureau of Economic Analysis, the Department of Commerce (BEA).

Commodity Prices. International Monetary Fund's index of world commodity prices. Source: International Financial Statistics.

Federal Funds Rate. Effective rate, monthly average. Source: Board of Governors of the Federal Reserve System.

GDP. Real GDP, seasonally adjusted, billions of chain 1992 dollars. Monthly real GDP is interpolated using the procedure described in Leeper, Sims, and Zha (1996). Source: BEA.

M2. M2 money stock, seasonally adjusted, billions of dollars. Source: Bureau of Labor Statistics (BLS).

Unemployment. Civilian unemployment rate (ages sixteen and over), seasonally adjusted. Source: BLS.

Because the model does not allow for judgmental adjustments periodically, it aims at strong performance of out-of-sample forecasting by the model itself (see Box 1 for details). When decision making is guided by forecasts extrapolated from the model, actual data for the future period are of course not available to policymakers. Therefore, out-of-sample forecasts, with probability distributions or error bands attached, can be invaluable. The error bands of forecasts give policymakers an indication of the range of the future data. Before the discussion turns to greater detail about the use of probability distributions of forecasts, the next three sections discuss out-of-sample point forecasts produced from the specific dynamic multivariate model presented here.

Out-of-Sample Point Forecasts

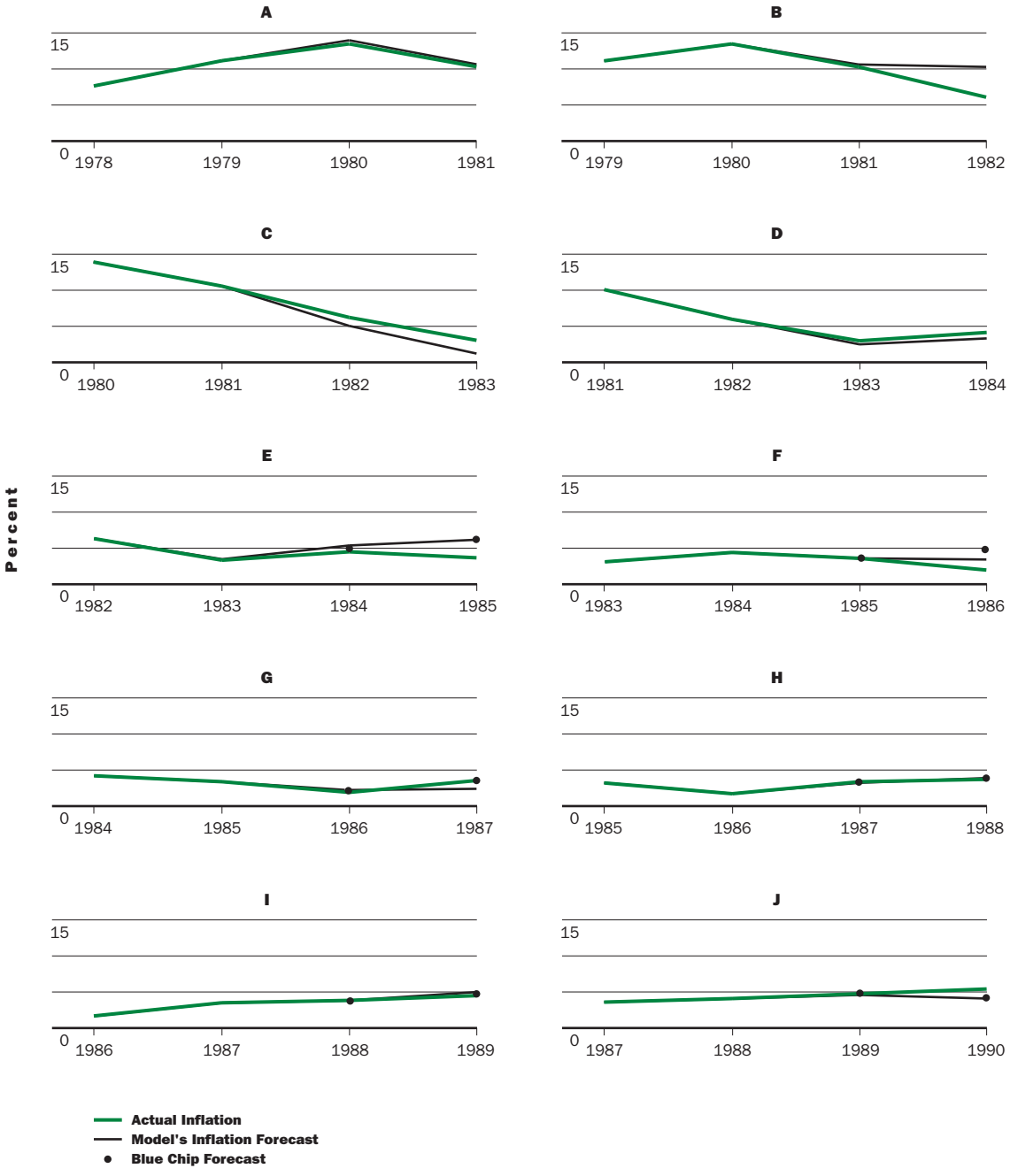
The 1980s. In the late 1970s inflation was accelerating to rates unprecedented in the period since 1960. Then in the 1980s inflation slowed down more quickly than the public thought possible. Thus, 1980s inflation is difficult to forecast. Chart 3 displays the model's forecasts of annual inflation through the 1980s. In each panel of Chart 3, the thick line represents actual outcomes of inflation, the thin line represents the model's forecasts for the next two years, and the dots are the Blue Chip forecasts for the next two years.[5] Note that the Blue Chip forecasts at the beginnings of 1980, 1981, 1982, and 1983 are not displayed here because the new methodology introduced to compute the CPI has significantly changed figures for actual inflation before 1984. New definitions or revisions of the data always affect the accuracy of evaluating the forecasts that were made using old data at the time. Inflation figures after 1983, however, have not been altered much by subsequent data revisions. In Panel E, for instance, the Blue Chip forecasts were made at the beginning of 1984. To be comparable, the model's forecasts are also made at the beginning of 1984. In addition, Panel E displays the actual data in the two years (1982 and 1983) prior to the forecast year. Similarly, in all other panels, the forecasts for the next two years are displayed along with the actual data in the two years prior to the forecast year. For example, in Panel F, inflation forecasts in 1985 and 1986 (the thin line and dots) are made at the beginning of 1985 along with actual inflation in 1983 and 1984.

As Chart 3 shows, without periodic judgmental adjustments the dynamic multivariate model here produces quite reasonable results that are as least as accurate as the Blue Chip forecasts. In particular, the model forecasts the slowdown of inflation in the 1980s fairly well. Because the model is dynamic, it adjusts its forecasts accordingly by systematically incorporating the most recent data. For example, at the beginning of 1981 the model tends to predict that the trend of inflation will be higher than that of actual outcome (Panel B); by the time 1981 is over, the model is able to predict the downturn of future inflation (Panel C).

How do the new data in 1981 help ameliorate the forecasting performance? Remember that the model is not only dynamic but also multivariate. The new data

4. *The mathematical structure is similar to Sims and Zha (1998). See Box 1 for details.*
5. Blue Chip Forecasts *is a monthly publication based on a survey of a number of forecasters from different industries. The Blue Chip forecasts displayed in this article are the consensus forecasts.*

CHART 3 Point Forecasts of Annual Inflation Rate, 1980s



Source: See Box 2; *Blue Chip Forecasts.*

include prices as well as the model's other macroeconomic variables, such as output, the interest rate, and the unemployment rate. The model systematically explores the dynamic relationships among these other variables and the CPI, complex though they might be. It is therefore unsurprising that inflation forecasting can be further improved by the model's ability to capture multivariate relationships in new data.

The 1990s. 1990 was a turning point for inflation. Since then, inflation has declined steadily, from 5.4 percent in 1990 to 2.9 percent in 1996. Such a favorable environment has, to a large extent, surprised the public and professional forecasters as well. Indeed, many forecasting firms have overpredicted inflation for this period. The 1990s is thus considered another very difficult inflation period to forecast. Nonetheless, the model's forecasts for this period, as shown in Chart 4, look reasonable in capturing the steadily declining pattern of inflation.

From Chart 4 one can see that since 1991, Blue Chip forecasts have been consistently higher than actual outcomes. The overprediction of inflation in the 1990s is consistent with simple rules of thumb such as the Phillips curve trade-off, given the declining unemployment rate after 1992. In contrast, the model's dynamic forecasts are more optimistic about the downward trend in inflation and closer to actual outcomes.

Regime Shifts. There is a common view that monetary policy follows simple rules and that these rules change from time to time in an exogenous fashion. For example, the 1979–82 period is often regarded as one in which the policy "rule" was completely changed because the Federal Reserve adopted new operating procedures to target nonborrowed reserves rather than the federal funds rate. After 1982 the Federal Reserve returned to targeting the federal funds rate. By this argument, the period after 1982 has been under a different regime than the 1979–82 period, and some empirical modelers use a sample period that begins only after 1982 as if the data before 1983 were irrelevant.

To examine this idea, the model here is reestimated using the data starting in 1983. Chart 5 reports inflation forecasts out of sample (indicated by the dots). Evidently, throwing away the data before 1983 does not improve out-of-sample forecasting in general and worsens it considerably in some cases (Panels D, E, and F).[6] One interpretation of these findings is that the Federal Reserve's behavior is complicated and cannot be characterized by discontinuous or abrupt changes in simple rules. Even among economists there is no agreement on whether the Federal Reserve's behavior during the 1979–82 period was actually different (Cook 1989). For example, Goodfriend (1993, 4) argues that "it is more accurate to refer to the period from October 1979 to October 1982 as one of aggressive federal funds rate targeting than one of nonborrowed reserve targeting." From a forecasting point of view, Charts 3 and 4 show that including data in this period helps forecast inflation in the 1980s and 1990s; Chart 5 suggests that in dismissing the data simply by a priori reasoning valuable information may be lost.

In a nutshell, the dynamic multivariate model that generates results in Charts 3–5 aims at accounting for both short-run dynamics and long-run relationships among the six key macroeconomic variables. Such a modeling strategy may explain the model's reasonable performance in forecasting inflation. Model-based forecasts provide benchmarks by which policymakers can decide on the best policy action given all current information. Furthermore, explicit modeling makes it easy to document the model's forecasting performance (as in Charts 3–5) and to continue improving the model or replace it by a better model when available.

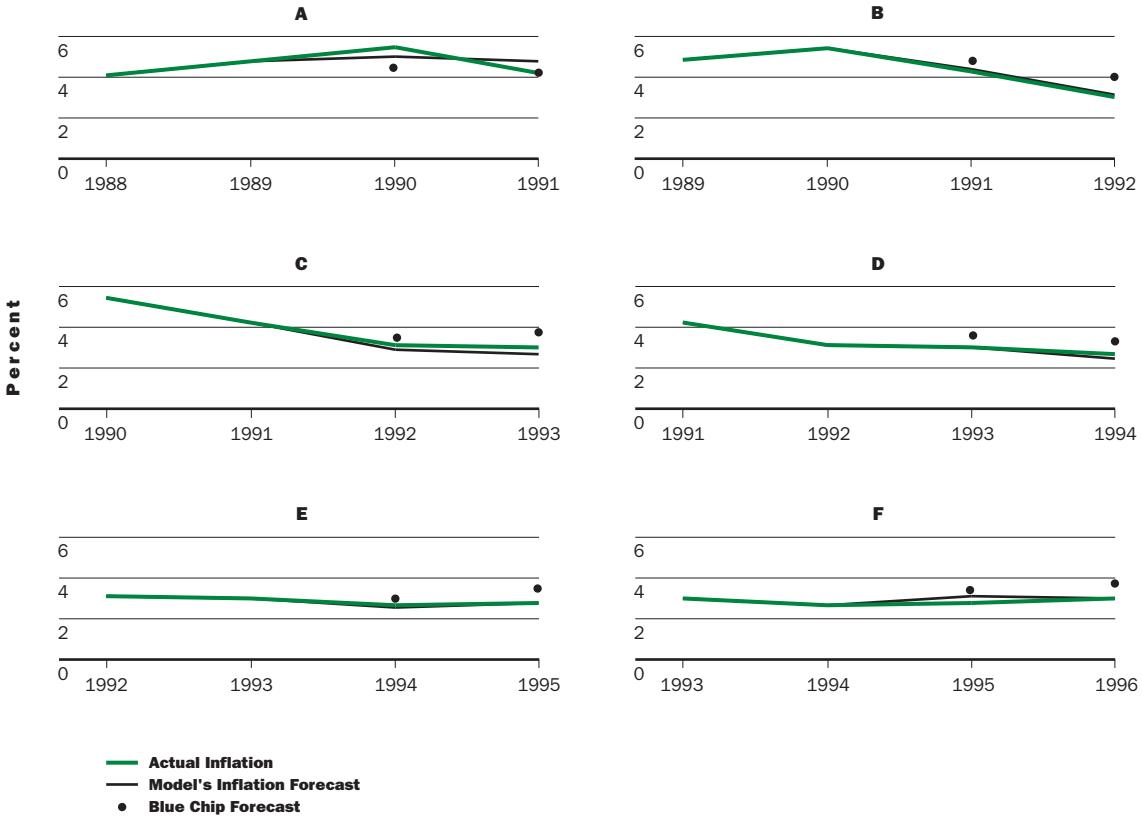
All models at best only approximate the actual economy. No model can forecast economic conditions with perfect accuracy. Thus, policymakers must use point forecasts cautiously and carefully.

The Distributions of Forecasts

All models at best only approximate the actual economy. No model can forecast economic conditions with perfect accuracy. Thus, policymakers must use point forecasts cautiously and carefully. When a model is used to advise policymakers, it is desirable that an explicit measure of uncertainty about the model's forecasts be provided. One effective way to measure uncertainty is to provide probability distributions of particular forecasts. With such a distribution, one is able to construct an error band on the forecast or to infer how likely the forecast is to be above or below a certain number. Error bands provide a sense of the uncertainty of economic conditions in the future and where the distribution of, say, inflation lies. Producing realistic error bands on forecasts has been a difficult technical problem. In a

6. *Technically, these two sets of forecasts may not be statistically different when error bands are considered. Small samples such as the data after 1982 tend to give unreliable results due to erratic sampling errors, as found in, say, Cecchetti (1995). The fact that the model with only the post-1982 data delivers reasonable results may be due to recent developments in Bayesian methods that deal with problems associated with low degrees of freedom (see Sims and Zha 1998 and also Box 1). This feature is still largely unexplored and deserves further research.*

CHART 4 Point Forecasts of Annual Inflation Rate, 1990s



Source: See Box 2; *Blue Chip Forecasts.*

recent paper Sims and Zha (1998) provide ways to compute probability distributions of forecasts from dynamic multivariate models (see also Box 1).

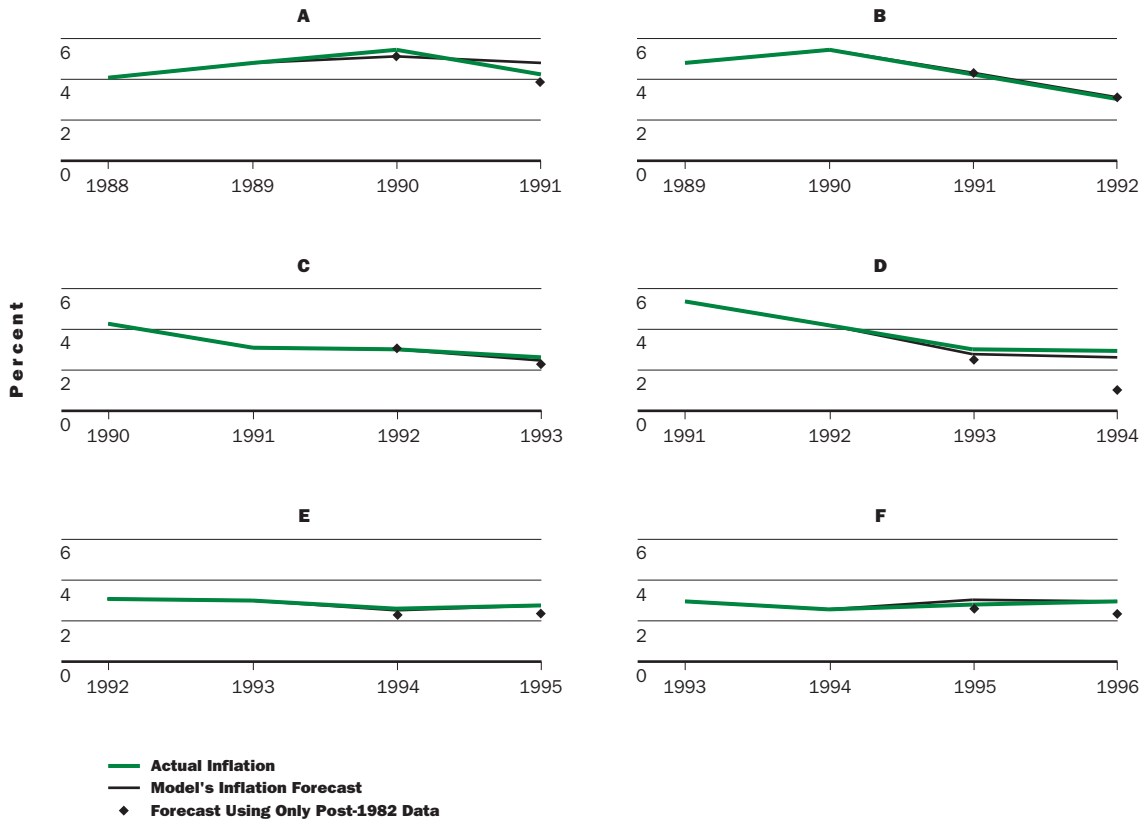
Given probability distributions of forecasts, error bands can be constructed for any desired probability. The purpose of constructing such a band is to demarcate reasonably high and low probability regions usable for policy deliberations. The error bands used in this article are constructed so that there is a two-thirds probability that the realized value is contained within the band. With this demarcation, events outside the band are given low probability and thus should be given less weight in decision making. One should bear in mind that low probability events do occur at times but less frequently.

As an example, Chart 6 presents the same forecasts as in Panel B of Chart 3 but with error bands attached. Whereas actual inflation for 1981 falls within the error band, actual 1982 inflation lies outside the error band. The error band at the two-year forecast horizon (that is, 1982) suggests that it is unlikely that 1982 inflation would return to the 1980 level, which indeed did not occur. At the same time, the model gives low probability to values far below 7.9 percent (the lower bound of the 1982 error band). But actual inflation in 1982 did occur at the level of 6.2 percent.

Most of the time, however, actual outcomes of inflation fall within error bands. This evidence is clear from Charts 3 and 4, in which point forecasts are often close to actual values of inflation. In addition to assessing quantitatively the uncertainty of forecasts, error bands provide ways of evaluating forecasts from other sources.[7] To show an example, Chart 7 displays the model's forecasts for the real GDP growth rate in 1995 and 1996 with error bands and Blue Chip predictions.[8] Actual GDP growth is inside the error bands, but the Blue Chip 1995 forecast of GDP growth at about 3.2 percent is far outside the error band. The model suggests that such a high growth rate is unlikely for 1995.

Although the error bands considered here are sufficient for most purposes, it is sometimes useful to know the entire distribution or likelihood that a particular forecast is going to be realized. Charts 8 and 9 provide two examples. Corresponding to Chart 6, Chart 8 presents the distribution of the inflation forecast for 1982. The two dashed vertical lines mark the band that contains two-thirds probability, and the solid vertical line marks the actual out-

CHART 5 Point Forecasts of Annual Inflation Rate, 1990s (Using post-1982 data)



Source: See Box 2.

come of inflation in 1982. The dispersed distribution in Chart 8 reflects a great uncertainty about inflation shortly after the high volatility of inflation during the late 1970s and early 1980s. Note that although actual inflation is outside the band, it is close to the lower bound of the band (that is, far away from the tail of the distribution).

Chart 9, corresponding to Chart 7, displays the distribution of the forecast of the real GDP growth rate in 1995. Again, the two dashed vertical lines mark the two-thirds probability band, the solid vertical line at 2 marks actual output growth in 1995, and the outer vertical line indicates the Blue Chip forecast. As can be seen in Chart 9, the Blue Chip forecast is near the tail of the distribution, implying that by the model's criterion such a forecast is very unlikely to be realized.

The discussion so far has been concerned exclusively with probability distributions or error bands around individual forecasts. While this focus is sufficient and effective for most policy analyses, it is important to bear in mind that individual forecasts are not independent of one another. Indeed, because of the multivariate nature of the model, forecasts of a set of variables of interest have a joint distribution. Such a distribution can be used to construct an error region that describes how likely forecasts of, say, both high output growth and low inflation are. Chart 10, for example, displays the error region that contains both real GDP growth and inflation for 1998 with a two-thirds probability.[9] The square represents the model's point forecast. The scattered circles are forecasts of real GDP growth and inflation for 1998 from fifty-five different

7. These sources can be various commercial firms, particular economic theories, institutional knowledge, or even ad hoc views.

8. All forecasts are made at the beginning of 1995. Although this article concentrates on inflation for simplicity of the analysis, forecasts of other macroeconomic variables such as output and unemployment are equally important for monetary policy. In particular, a number of economists believe that there is a short-term trade-off between inflation and output, especially when unexpected large shocks hit the economy (King 1997).

9. Similar to error bands of individual forecasts, error regions of joint forecasts can be constructed for any desired probability. Again, the discussion here focuses on two-thirds probability.

CHART 6 Inflation Forecasts with Error Bands for 1981 and 1982

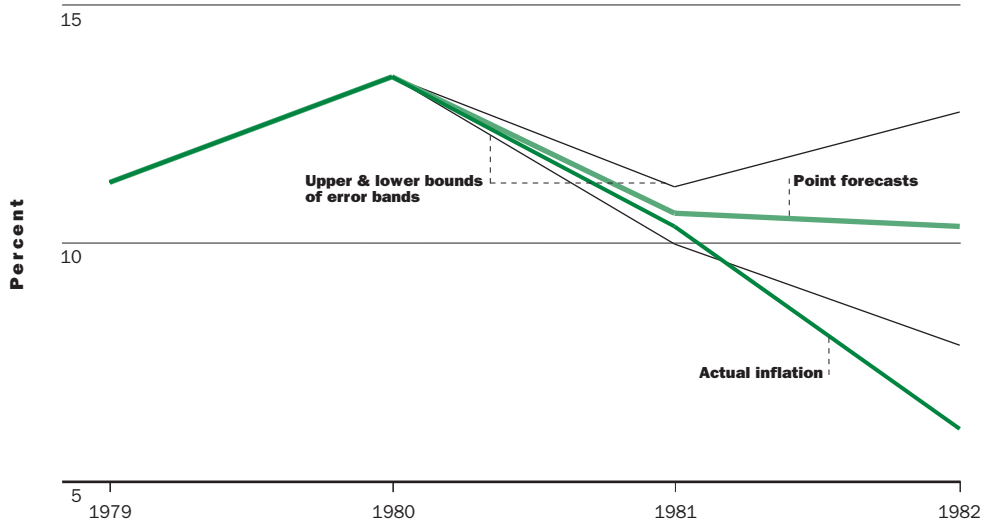
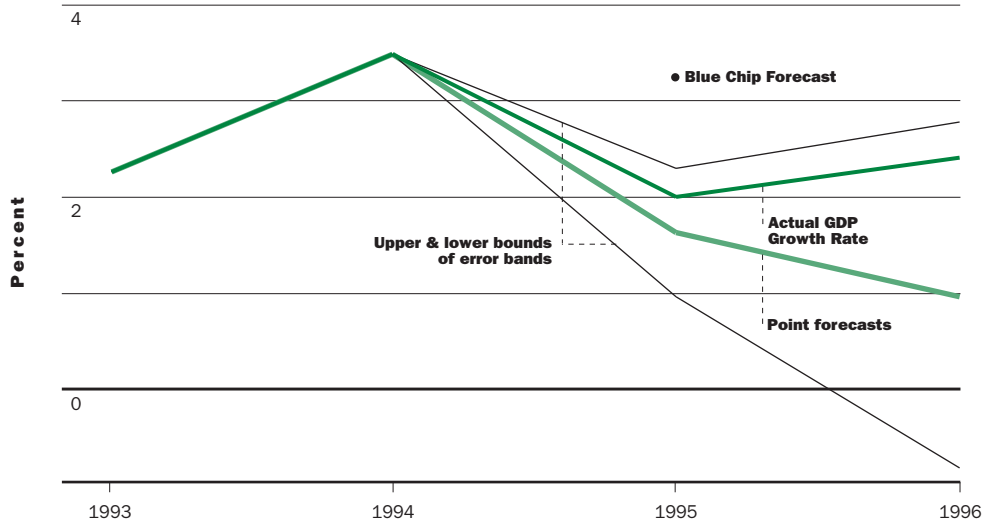


CHART 7 Real GDP Forecasts with Error Bands for 1995 and 1996



firms, published by the *Wall Street Journal* on January 2, 1998. Because these forecasts were submitted by December 18, 1997, the model's 1998 forecasts and error region in Chart 10 were made as of December 1997 to be as compatible with the *Wall Street Journal* forecasts as possible.[10] According to the error region, the model gives as much probability to the scenario of high GDP growth (3.5–5.5 percent) and low inflation (around 2 percent) as to that of medium GDP growth (2–3.5 percent) and low inflation (around 2 percent). But the model gives low probability to the scenario of low GDP growth (under 2 percent). The *Wall Street Journal* forecasts are unequally dispersed. At least one-fifth of the firms produced forecasts outside the model's error region. None of the firms produced forecasts that fall within the top half of the error region implied by the model.

Conclusion

The real world of monetary policy is complex. Because of long and variable lags in the effects of policy actions, the Federal Reserve faces a difficult task in trying to achieve its multiple objectives. The

CHART 8 Distribution of Inflation Rate Forecast for 1982

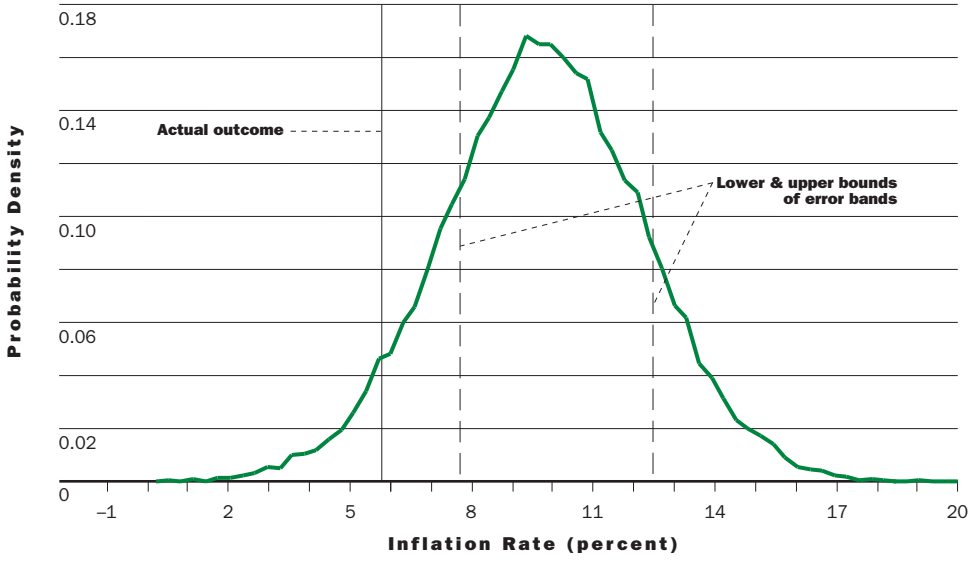
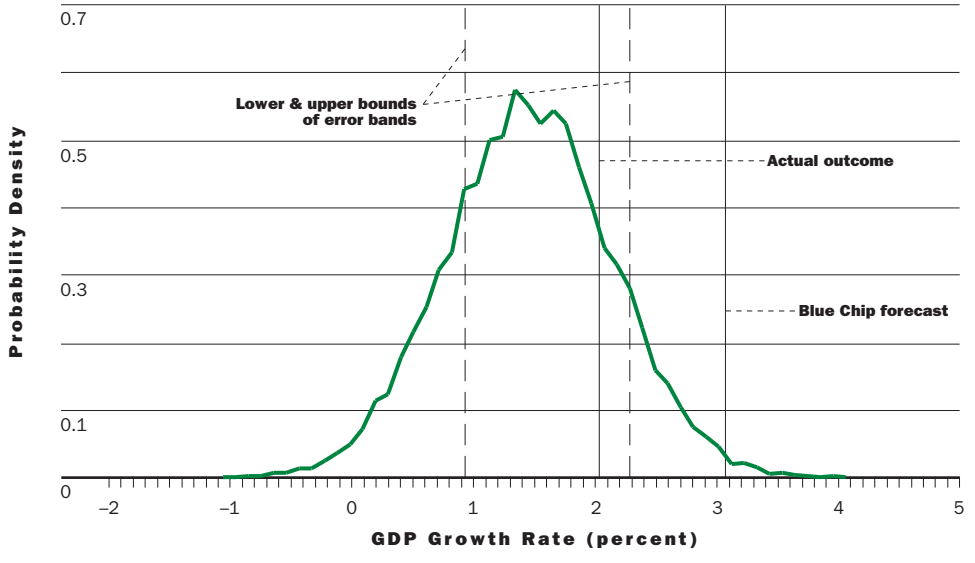


CHART 9 Forecast Distribution of GDP Growth Rate for 1995

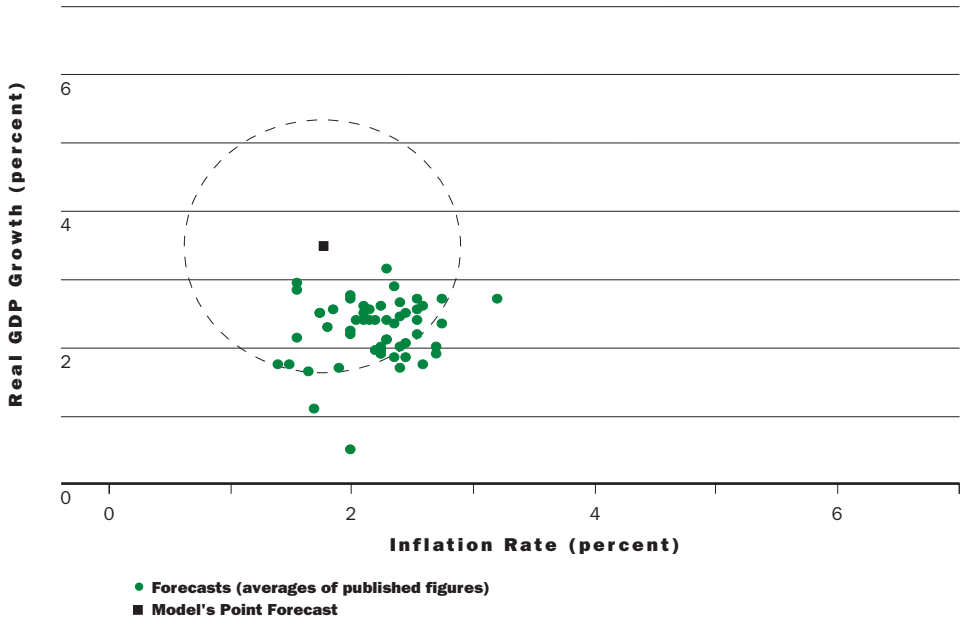


foregoing discussion concentrates on only one of these objectives—to keep the path of inflation low and stable. Given this objective, policy projections under different paths of a policy instrument (for example, the federal funds rate) are an integrated part of forward-looking policy formation. And reliable forecasts of the path of inflation are the first step in this process (Bernanke and Mishkin 1997).

The dynamic multivariate model discussed here is transparent enough to be reproduced and improved. At the same time, it is sufficiently complex to capture dynamic interplay between policymakers and the private sector. Consequently, it shows reasonable performance in forecasting as compared with other forecasts. More important, this approach provides empirically coherent ways to assess the uncertainty inherent in forecasts. Error

10. The forecasts displayed in Chart 10 are the 1998 averages of published figures in the Wall Street Journal.

CHART 10 Error Region for Forecasts of Real GDP Growth and Inflation Rates for 1998



Source: *Wall Street Journal,* January 2, 1998.

bands or distributions of forecasts are essential for gauging this uncertainty in at least two aspects. First, they offer an assessment of how likely or realistic other forecasts are. Second, error bands inform policymakers of the uncertainty they face, reminding them of the "need to be flexible in revising forecasts and the policy stance in response to new information contradicting their previous predictions" (Kohn 1995, 233).

As Chairman Greenspan has observed, "Operating on uncertain forecasts, of course, is not unusual. . . . [I]n conducting monetary policy the Federal Reserve needs constantly to look down the road to gauge the future risks to the economy and act accordingly" (1997b, 17). The dynamic multivariate model presented in this article provides a useful tool for gauging future uncertainty and an empirically consistent way to update forecasts. It is hoped that future research will apply such a model to tasks of policy projections.

REFERENCES

Bernanke, Ben S., and Frederic S. Mishkin. 1997. "Inflation Targeting: A New Framework for Monetary Policy?" *Journal of Economic Perspectives* 11 (Spring): 97–116.

Blinder, Alan S. 1997. "What Central Bankers Could Learn from Academics—and Vice Versa." *Journal of Economic Perspectives* 11 (Spring): 3–19.

Board of Governors of the Federal Reserve System. 1994. *The Federal Reserve System: Purposes and Functions.* Washington, D.C.

Cecchetti, Stephen G. 1995. "Inflation Indicators and Inflation Policy." *NBER Macro Annual 1995,* 189–219.

Chang, Roberto. 1997. "Is Low Unemployment Inflationary?" Federal Reserve Bank of Atlanta *Economic Review* 82 (First Quarter): 4–13.

Christofferson, Peter F., and Francis X. Diebold. 1997. "Cointegration and Long-Horizon Forecasting." *Journal of Business and Economic Statistics*, forthcoming.

Cook, Timothy. 1989. "Determinants of the Federal Funds Rate: 1979–1982." Federal Reserve Bank of Richmond *Economic Review* 75 (January/February): 3–19.

Diebold, Francis X. 1998a. "The Past, Present, and Future of Macroeconomic Forecasting." *Journal of Economic Perspectives,* forthcoming.

———. 1998b. *Elements of Forecasting.* Cincinnati, Ohio: South-Western College Publishing.

Espinosa, Marco A., and Steven Russell. 1997. "History and Theory of the NAIRU: A Critical Review." Federal Reserve Bank of Atlanta *Economic Review* 82 (Second Quarter): 4–25.

Friedman, Milton. 1992. *Money Mischief: Episodes in Monetary History.* New York: Harcourt Brace Jovanovich.

Goodfriend, Marvin. 1993. "Interest Rate Policy and the Inflation Scare Problem: 1979–1992." Federal Reserve Bank of Richmond *Economic Quarterly* 79 (Winter): 1–24.

Greenspan, Alan. 1997a. *Statement before the Senate Committee on Banking, Housing, and Urban Affairs.* February 26.

———. 1997b. *Statement before the House Subcommittee on Domestic and International Monetary Policy, Committe on Banking and Financial Services.* July 22.

King, Mervyn. 1997. "The Inflation Target Five Years On." *Bank of England Quarterly Bulletin* (November): 434–42.

Kohn, Donald L. 1995. "Comment on 'Inflation Indicators and Inflation Policy' by Cecchetti." *NBER Macro Annual 1995,* 227–35.

Leeper, Eric M., Christopher A. Sims, and Tao Zha. 1996. "What Does Monetary Policy Do?" *Brookings Papers on Economic Activity* 2:1–63.

Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica* 48 (January): 1–48.

Sims, Christopher A., and Tao Zha. 1996. "Does Monetary Policy Generate Recessions?" Yale University and Federal Reserve Bank of Atlanta, manuscript.

———. 1998. "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review,* forthcoming.

Staiger, Douglas, James Stock, and Mark Watson. 1997. "The NAIRU, Unemployment, and Monetary Policy." *Journal of Economic Perspectives* 11 (Winter): 33–50.

Wessel, David. 1997. "Fed Lifts Key Short-Term Interest Rate Amid Speculation More Rises Are Likely." *Wall Street Journal,* March 26.

The Rise of Risk Management

J. DAVID CUMMINS, RICHARD D. PHILLIPS, AND STEPHEN D. SMITH

Cummins is the Harry J. Loman Professor of Insurance and Risk Management at the Wharton School, University of Pennsylvania, and a senior fellow at the Wharton Financial Institutions Center. Phillips is an assistant professor at Georgia State University. Smith holds the H. Talmage Dobbs Jr. Chair of Finance, Georgia State University, and is a visiting scholar at the Atlanta Fed.

The authors thank Robert Bliss, Gerald Dwyer, Saikat Nandi, and Mary Rosenbaum for helpful comments.

RISK MANAGEMENT CAN BE ROUGHLY DEFINED AS ANY SET OF ACTIONS TAKEN BY INDIVIDUALS OR CORPORATIONS IN AN EFFORT TO ALTER THE RISK ARISING FROM THEIR PRIMARY LINE(S) OF BUSINESS. VIEWED FROM THIS PERSPECTIVE, RISK MANAGEMENT IS NOTHING NEW, DESPITE THE INCREASED ATTENTION GIVEN TO THE SUBJECT BY ACADEMICS AND MANAGERS AS FINANCIAL DERIVATIVE MARKETS HAVE EVOLVED OVER THE PAST DECADE OR TWO.

For well over one hundred years farmers, for example, have engaged in risk management as they attempted to hedge their risks against price fluctuations in commodity markets. Their preferred risk-management strategy has been to sell some or all of their anticipated crop, before harvest time, to another party on what is called futures markets. This strategy guarantees the farmer a known price for his crop, regardless of what the commodity's actual price turns out to be when the harvest comes in. Risk management along these lines makes sense for farmers for at least two reasons. First, agricultural prices are volatile. Moreover, many of these family farmers are not diversified and, in addition, must borrow in order to finance their crops. Therefore, setting the sale price now shifts their risk of price fluctuations to other participants in the futures market better able or willing to bear this volatility.

Contrast the above story with that of a large corporation, owned by a large number of shareholders, facing similar commodity price risk. For concreteness, consider a firm primarily engaged in the extraction and sale of copper. Given that copper prices are relatively volatile, the first rationale for risk management might seem similar to the farmer's. However, unlike in the farmer's circumstance, this firm is owned by a large number of shareholders, who can, if they so wish, greatly reduce or eliminate the risk that copper prices will be low simply by holding a diversified portfolio that includes only a small fraction of assets invested in the copper extraction corporation. More generally, if investors can freely trade securities in many firms, they can choose their exposure to volatility in copper prices. Indeed, in two studies Modigliani and Miller (1958; Miller and Modigliani 1961) showed that, in a world with no transactions costs or taxes and with equal information, managers could not benefit their shareholders by altering the risk profile of the firm's cash flows. Essentially, in this situation shareholders can already do whatever they choose at no costs; actions by managers are redundant.

Although the Modigliani and Miller studies considered the option of changing the firm's risk profile only through the use of debt financing (1958) or the distribution (or lack thereof) of dividends (Miller and Modigliani 1961) and not through the use of financial derivative securities, the powerful intuition here is the same as that outlined earlier. If managers of the firm wished to increase their use of debt financing (say, because they thought it was cheaper than equity), investors could undo this transaction by, for example, taking equal positions in the firm's debt and equity. This move would leave investors facing the same risk from the firm's operations as they had before the increase in debt financing.

Given the above discussion, one is tempted to ask, Why are managers doing for shareholders what shareholders apparently can do for themselves? In other words, why do managers of corporations find it worthwhile to engage in risk-management activities, and why has interest in this topic mushroomed over the past decade or two?

This article is intended to provide a review of the rationales concerning when firms should engage in risk management. The first section lays the groundwork for that discussion by defining more precisely what risk management is in terms of the alternative instruments available to managers. The second section provides a discussion of the reasons firms might have for managing risk, while the third section summarizes the empirical evidence concerning the actual economic factors associated with using one such set of instruments, namely, derivative securities. Since derivatives exist solely for purposes of managing risk, studies of this type are relatively "clean" tests of the various rationales put forth for why corporations manage risk.

What Constitutes Risk-Management Activities?

The term *risk management*, at its most general level, simply denotes a situation in which an individual or firm makes decisions to alter the risk/return profile of future cash flows. The terminology typically used is that if managers are attempting to reduce risk through their actions, they are said to be hedging; if managers are trying to increase the firm's risk exposure because they believe that such a strategy will yield abnormal profits, they are said to be speculating.

To put the decision about engaging in risk management in some perspective, this section of the article outlines the types of activities most commonly thought of as risk management. For concreteness, again consider the problem faced by the firm engaged in copper extraction. Because commodities markets are competitive in the sense that one firm's activities will typically have a very small impact on market prices, the underlying risk—copper price fluctuations—facing this firm can be generally seen as given.

How might the firm alter this risk?[1] One approach would involve diversifying its product line. That is, management could divert some of the firm's resources to the extraction of some other commodity—silver, perhaps—and to the extent that copper and silver prices do not move in perfect unison, doing so would lower the firm's net risk.

Secondly, the firm could try to manage its expenditures so that they would tend to increase when revenues are high and fall when copper prices (and sales dollars) are abnormally low. For example, the firm could shift extraction methods away from those relying heavily on capital assets (with their fixed costs) to those methods depending more on labor or other inputs that would be viewed as variable costs. Under this scenario, when copper prices increase the firm can hire more workers and when prices fall unusually low they can lay off some of the workforce. In this situation, fluctuations in investors' net income are less than if the firm uses a more automated technology, which requires payments on the machines whether copper prices turn out to be high or low. Thus, changes in operating leverage could be viewed as a form of risk management.

A third possibility would be for the firm to reduce its leverage—its percentage of financial capital raised through the sale of debt securities. In this case, fluctuations on the firm's return on invested capital result in smaller fluctuations in the return-to-equity capital. In short, the firm's choice of debt versus equity financing can be viewed as a form of risk management.

Another way that management can alter the distribution of cash flows involves the use of derivative securities, so named because their price depends on the price of some underlying instrument (such as stocks or interest rates). While modern derivatives contracts can be, in many ways, exceedingly complex, almost all these types of instruments essentially consist of some combination of options and forward contracts. Moreover, the claims are linked, in the sense that one can, for example, replicate the cash flows from a forward contract by simultaneously buying certain options and selling others.[2] Options are contracts that, for an up-front fee, give the purchaser the opportunity, over some period of time, to buy or sell something (for example, a share of

1. *The examples discussed here focus on reducing risk. Of course, if managers wanted to increase the risk faced by a firm's shareholders, they could reverse these actions.*
2. *Cox and Rubinstein provide an excellent discussion of options and a detailed analysis of how the prices of these securities are determined as a function of the prices of underlying securities. They also provide a concise treatment of the cash flow replication idea discussed in the text (1985, 59–60).*

If Derivatives Are Used to Hedge Risk, Why Do Some Firms Lose So Much Money?

Is there a significant downside to the use of derivatives in risk management? A casual glance at press reports over the past few years seems to indicate that derivatives are excessively risky and, in fact, dangerous to the financial health of corporations and other derivatives traders. Among the most widely publicized derivatives debacles are the losses of Proctor and Gamble ($137 million) and Gibson Greetings ($20 million) as a result of transactions in interest rate swaps.[1] Orange County, California, and the Orange County Investment Pool (OCIP) declared bankruptcy in 1994 following a $1.7 billion drop in the market value of the pool due to transactions in leveraged intermediate-term fixed-income securities. And perhaps the most spectacular example is the 1995 collapse of Barings Bank, a highly respected British merchant bank, due to losses of $1.3 billion on options and futures transactions in the Japanese stock and bond markets. Barings financed the U.S. Louisiana Purchase from France in the early nineteenth century and was banker to the royal family. The collapse of such a historically significant financial institution was all the more surprising given that it reported record profits for 1994.

These examples reveal the truism that derivatives, like other risky securities, can expose traders to the risk of substantial losses. However, the proper conclusion to be drawn from these cases is not that derivatives should be avoided but rather that participants should have the expertise and oversight systems that would be common for other investment and trading activities. In each of the cases cited above, the losses can be traced to inappropriate behavior on the part of one or more parties involved in the derivatives transactions. Some authors (see, for example, Smith 1997) argue that, at least in the case of Proctor and Gamble, lack of expertise by managers in assessing market risk seems to have played a role. However, both Proctor and Gamble and Gibson Greetings collected substantial damages from the counterparty in their interest rate swap contracts (Bankers Trust Company). The Securities and Exchange Commission (SEC) concluded that Bankers Trust had defrauded Gibson Greetings, and the Proctor and Gamble case was settled out of court. The SEC later cited Gibson Greetings for failing to disclose properly its derivatives-related profits and losses. The company was also sanctioned for having inadequate internal controls to ensure that its derivatives transactions were accounted for in accordance with generally accepted accounting principles. In the case of Orange County, the investment manager for the investment pool entered into inappropriate speculative transactions.[2] And, in the Barings Bank case, inadequate supervision and controls allowed a rogue trader to run up millions of dollars in losses while concealing his positions from superiors.

The message from recent derivatives debacles thus seems clear: derivatives positions need to be carefully designed and managed and controls should be in place to ensure that positions taken are fully understood by and consistent with the objectives of the organization. The need for expert management and control is the source of most of the fixed costs of entering derivatives markets, discussed in the text. That is, organizations planning to enter derivatives markets must put in place a team of investment managers who can structure an effective derivatives program, and monitoring and control systems must be created that prevent fraud and mismanagement. Of course, the same cautionary message applies to other activities undertaken by firms that involve substantial sums, such as investments in new projects, capital structure decisions, and mergers and acquisitions. In this regard, derivatives are not really different from other transactions conducted by firms. They are simply newer, and therefore many firms have acquired less experience in their management or have failed to implement appropriate accounting and control systems. The message from the derivatives debacles is that firms should acquire the appropriate human expertise in the areas of both trading and control before entering the market.

1. *For more on the Proctor and Gamble and Gibson Greeting cases see Smith (1997) and Overdahl and Schachter (1995), respectively.*
2. *There is also evidence that mismanagement after the decline in the market value of OCIP exacerbated Orange County's losses. Miller and Ross (1997) suggest that OCIP was neither insolvent nor illiquid in December 1994. They argue that OCIP should not have been liquidated and that the suspect financial instruments should have been held to maturity. This strategy would have enabled the county to avoid some of its losses and realize substantial net cash inflows during 1995.*

stock), with the sale price fixed today. A forward contract is an agreement between two parties to engage in a trade at some point in the future, with the terms of trade (for example, the sale price) set today.[3]

Although derivatives problems have made the news in recent times (see Box 1), they are no more or less risk-management tools than the other available alternatives discussed above. Indeed, Peterson and Thiagarajan (1997), among others, have argued that one cannot meaningfully assess whether one firm is more or less engaged in risk management without knowledge of all important operating, financial, and accounting decisions. They find evidence in their case studies to suggest that some managers use accounting practices that tend to smooth earnings, along with decisions concerning operating and financial leverage, as substitutes for trading in derivative securities.

Recent Advances in the Theory of Risk Management

The puzzle the introduction outlined involves the question of why managers of widely held corporations, acting in the interest of their stockholders, should manage risk that their shareholders could presumably manage themselves. Given the nature of this statement, the answer must, roughly speaking, lie in one of two areas: either there are some risks that shareholders cannot manage for themselves as inexpensively or managers are acting in their own interests, rather than those of the stockholders of the firm. There are proponents for each of these points of view, as discussed below.

Managerial Motives for Risk Management. Managers themselves may engage in risk-management activities because they have disproportionately large investments (their skills or human capital) in the firm they manage and, unlike shareholders, cannot easily diversify this personal risk. Being averse to risk, they are concerned about negative shocks to profits, particularly those that might bring the firm to the brink of bankruptcy. Bankruptcy or, more generally, times of financial distress are often associated with the replacement of current management. Thus, these undiversified managers are in much the same position as the farmers discussed in the introduction, and they might well be willing to engage in risk-management practices that will generate positive cash flows should the firm fall on bad times, at the cost of reducing cash flows in the good times.

Consider again the firm primarily engaged in the extraction of copper. According to traditional finance theory (for example, Sharpe 1964), shareholders care only about the systematic risk of their holdings, that is, only that risk that cannot be eliminated by having small investments in many different types of firms. Given that hedging copper prices may be costly in terms of lower average future income (after all, insurance is not typically free), stockholders would not be inclined to support actions by management that reduce risk that is viewed as diversifiable; namely, they would not share management's consternation about the financial difficulties or even the failure of one particular corporation. Smith and Stulz (1985) provide formal discussions of this issue. It is also intuitively clear why the manager would favor such activities—job protection. To the extent that managers have an excess investment in human capital in the firm and it is costly to transfer these skills should they need to seek other work, they have an economic incentive to have the firm continue as a going concern.

While modern derivatives contracts can be, in many ways, exceedingly complex, almost all these types of instruments essentially consist of some combination of options and forward contracts.

Shareholders may tolerate such potentially value-reducing activities if their managers are viewed as having other unique value-enhancing skills, bankruptcy is not costless, managers demand higher compensation in return for the risk they face, or confronting management is costly in terms of time and effort. Individual shareholders with, by design, relatively small stakes involved in a given firm may simply attempt to "free ride" and hope that some other group of stockholders will take up the cause of replacing management. But, of course, the other shareholders may be thinking the same thing, and often no action is taken.[4]

Rationales for Risk Management that Enhances Value. Numerous reasons have been put forth to argue that it really may be in shareholders' interests for certain types of enterprises to manage risk. The following is an incomplete sampling of the specific rationales, but the two general points are that there may be some risks

3. In some cases the terms of trade allow one or another of the parties some latitude concerning, for example, what exactly will be exchanged.

4. Of course, if things get bad enough in terms of too many value-reducing activities, outsiders with large amounts of capital may try to take over the firm, for example, by offering to buy up the shares of the firm's stockholders. Grossman and Hart (1981) note that there is a free rider problem here as well ("I will not sell my shares now. Rather, I will hold my shares until the new management improves firm performance and then sell at a profit.").

that are not tradable and that there exist situations in which there are informational differences among owners and managers. The existence of nontradable risk limits the degree of homemade diversification that shareholders can achieve; managing these risks is *not* something the shareholders can do for themselves. Informational differences can result in undervaluation of some firms, which is clearly not in the interests of the corporation's shareholders.

What are some of the noninformational frictions that might lead to a demand for risk management? First, whatever the underlying, value-related motive for risk management, the existence of fixed costs associated with using derivative instruments may make it more likely that only larger firms, with the resources to pay these large up-front costs, will manage risk through derivatives trading.[5] Second, if bankruptcy or financial distress imposes costs on the firm, shareholders may be willing to hedge profits in an effort to forgo these costs (see for example, Smith and Stulz 1985). These costs include both the direct legal and regulatory costs of bankruptcy as well as the indirect costs resulting from deteriorating relationships with key employees, suppliers, or customers. The indirect costs can have an adverse impact on the firm's cash flows even in the event bankruptcy is not the ultimate outcome. This dynamic suggests that firms with more fixed obligations—for example, debt obligations—will be willing to hedge more, other things held constant (see Brennan and Schwartz 1988).

It is also the case that many tax write-offs, such as depreciation, are not independently tradable, although they may be carried forward. However, given the time value of money, it may make sense for the firm to hedge against situations (for example, extremely low copper prices) in which it cannot exploit its tax deductions because income is low or negative.[6] Furthermore, the very fact that, other things held constant, corporate taxes are increasing at a nondecreasing rate in before-tax corporate profits provides another potential motivation for hedging. Smith and Stulz (1985) show that the firm can minimize its expected tax bill by keeping the volatility of income low (staying in the middle of the tax schedule). For example, given today's corporate tax code (and ignoring the alternative minimum tax), a firm with a fifty-fifty chance of having taxable income of $70,000 or $0 will have an expected tax bill of $6,250 while one with a sure taxable income of $35,000 will pay a tax of $5,250, an expected tax savings of $1,000.[7] While this factor might appear to be unimportant for most corporations (the marginal tax rate flattens out at taxable income of around $18,000,000), Graham and Smith (1996) provide evidence that, because of factors such as tax-loss carry provisions, tax effects may be more pronounced than would appear at first glance, especially for firms whose before-tax incomes tend to fluctuate between large positive and negative values.

In both of the above cases, shareholders might rationally support managers in their attempt to moderate income fluctuations by using risk-management tools, such as locking in at least some component of future income by being short forwards or futures contracts in copper or reducing fixed costs so that there is less fluctuation in pretax income. Using the same reasoning as above, firms that finance themselves with generally illiquid, if not outright nontradable, debt securities (for example, privately placed bonds) or hold particularly illiquid assets (such as collateralized mortgage obligations with unconventional repayment schedules) might find hedging their fluctuations in income or value worthwhile.

One might be tempted to ask why, if a firm is fundamentally sound but in temporary distress, managers do not simply keep these assets and liabilities on the books and raise additional outside funds. Froot, Scharfstein, and Stein (1993) argue that in a world of differential information between managers and potential outside investors, firms may encounter situations in which funds are needed but outside capital either is not available or is too costly. In such a case, managers may increase the current value of their firms by entering into contracts (the example they use is forward contracts) that generate positive payoffs when the firms' cash flows from operations turn out to be low.

The essence of the argument by Froot, Scharfstein, and Stein and others is that if there is asymmetric information between those who manage the firm and outside investors, better-than-average firms will have to sell securities to outsiders at a discount (less than the full-information value of the claims on the firm). By engaging in risk-management activities, these firms can avoid having to go to capital markets to acquire funds during a period of temporarily poor performance. This follows from the fact that their risk-management contracts are designed to pay off when the firm is otherwise doing poorly.[8] Notice, however, that if the firm keeps relatively large cash balances, there is less need to worry about times when the company is "short cash," and one would therefore expect larger levels of liquidity to be associated with less risk-management activity.

What Kinds of Firms Manage Risk and Why Do They Do It?

As mentioned earlier, managers can manage risk using a wide variety of tools. However, unlike some traditional methods, like changing operating or financial leverage, derivative securities exist only for purposes of risk management. Tests with these data therefore provide somewhat "cleaner" results concerning why firms may choose to engage in risk management.[9] It is also the case that the volume of activity in

derivatives contracts has grown dramatically over the past two decades. Box 2 provides some details on the overall growth of derivatives transactions and some summary data concerning what firms are actually engaging in these transactions. With these points in mind, this section first provides a review of results from some recent empirical studies that test the primary theoretical hypotheses relating to why firms actually do (or do not) use derivative securities for risk management.

Why Firms Manage Risk. A major study investigating the question of motive is by Tufano (1996), who looks at managerial compensation schemes and hedge ratios in the gold mining industry in an attempt to contrast managerial motives with those associated with value-maximizing theories of risk management. Hedge ratios are usually defined as the percentage of expected future production that the firm has effectively sold short through risk-management activities: in this case the firm is using derivative securities like short futures positions (that is, agreeing to sell gold forward with the price fixed today) or the purchase of put options (purchasing the right to sell gold in the future at a price fixed today) on gold.

Tufano argues that risk-averse managers whose compensation comes in large part through acquiring shares in the firm will want to hedge their risk. As discussed at length earlier, such a policy would not necessarily benefit diversified shareholders, so, to the extent that there are costs associated with hedging, the manager is better off and the shareholders worse (or at least no better) off than if the firm abstained from risk management altogether. He contrasts these managers with those who earn a relatively large portion of their compensation through the granting of stock options (call option contracts on the stock of the firm). In this situation managers can walk away from the options should the firm do poorly, but if the firm does well their positions will provide high payoffs. In a "heads I win, tails you lose" environment like this, even risk-averse managers would be more willing to tolerate gold price, and therefore earnings, fluctuations. Thus, they would find it less advantageous to hedge.

The evidence from studies investigating the decision by financial companies to use derivatives as a way to avoid financial distress costs is mixed.

Tufano finds support for this hypothesis in the data. In particular, his evidence suggests that managers with high option holdings manage risk less than those with high stock holdings. Such results are consistent with the managerial risk-aversion hypothesis of risk management.[10] Tufano claims to find almost no evidence in favor of the various rationales that would make risk management a value-maximizing decision and thus in the interests of shareholders. He does find, however, that firms with large cash balances tend to manage risk less. This finding is consistent with the hypothesis that firms with less risk of having to seek outside financing, other things being the same, will hedge less.[11]

Contrary to Tufano's results, some authors have provided evidence that they believe is consistent with value-maximization theories of risk management. The

5. *However, there is an offsetting notion that suggests that larger firms have more built-in diversification (more independent product lines) and therefore should have less need for the services provided by risk management. This hypothesis would be the alternative associated with the fixed-cost idea discussed in the text.*
6. *MacMinn (1987) provides a rigorous analysis of this issue.*
7. *These calculations follow from the fact that the first $50,000 in taxable income is taxed at the rate of 15 percent, while the income between $50,000 and $70,000 is taxed at a rate of 25 percent.*
8. *Froot, Scharfstein, and Stein argue that another condition needed for this type of hedging to be valuable is that the firm's production function display decreasing returns to scale—that is, the firm's output is increasing in its inputs but at a decreasing rate. Alternative assumptions can substitute for this condition. For example, any type of asset that is indivisible (for example, it is difficult to sell one office in an office building), when combined with the scenario of asymmetric information, will do the job in the sense that a firm might rationally want to hedge against the possibility that they might end up having to sell all this valuable asset at an unfavorable price when they only need a small amount of cash to pay creditors or invest in some new growth opportunity.*
9. *However, even these are not perfectly unambiguous tests since it is difficult to control for all of the other risk-altering strategies undertaken by the managers of these firms.*
10. *As a substitute for these actions, managers could hedge their risk on their own personal accounts. However, effectively hedging the risk of adverse movements in the stock of the firm may require, for example, managers to short sell the stock of the firm for which they work. This, or economically similar actions such as buying puts on the firm's stock, may be contractually prohibited or, at a minimum, send a bad signal to outside shareholders. Therefore, managers may choose to avoid hedging on personal accounts. Moreover, even if trading in, say, gold futures may not be prohibited, it is still the case that transacting at the firm level spreads the transactions costs across all shareholders.*
11. *Mian (1996), using a larger set of industrial firms, finds a similar negative relationship between the level of liquid balances and the degree of risk-management activities on the part of corporations.*

The Growth of Derivatives

The worldwide derivatives market has grown dramatically in the last decade to become a significant component of the world's financial markets. A 1996 survey, conducted by twenty-six central banks, estimated the worldwide volume of derivative contracts, measured as the total notional value of derivative contracts outstanding,[1] to be approximately $55.7 trillion as of the end of March 1995 (Bank for International Settlements 1995). The market value of the potential cash flows from these contracts was estimated to be approximately $2.2 trillion.

To get a better idea of how fast the market for derivatives has grown over the last decade, consider the chart, which displays the total worldwide notional value of all privately negotiated interest rate swap, interest rate option (including caps, floors, collars, and swaptions), and currency swap contracts outstanding over the period from 1987 to 1996. Based on the data in the chart, the average yearly growth rate of the notional value of these contracts is more than 40 percent annually. Likewise, in the United States, the growth rate of derivative transactions by banks, insurers, and securities firms also has been impressive. Although not quite as high as the worldwide rate, the average annual growth rate of the notional value of derivative contracts outstanding over the 1990–95 period for the fifteen largest over-the-counter derivatives dealers in the United States was 27 percent, as reported by the U.S. General Accounting Office (1997).

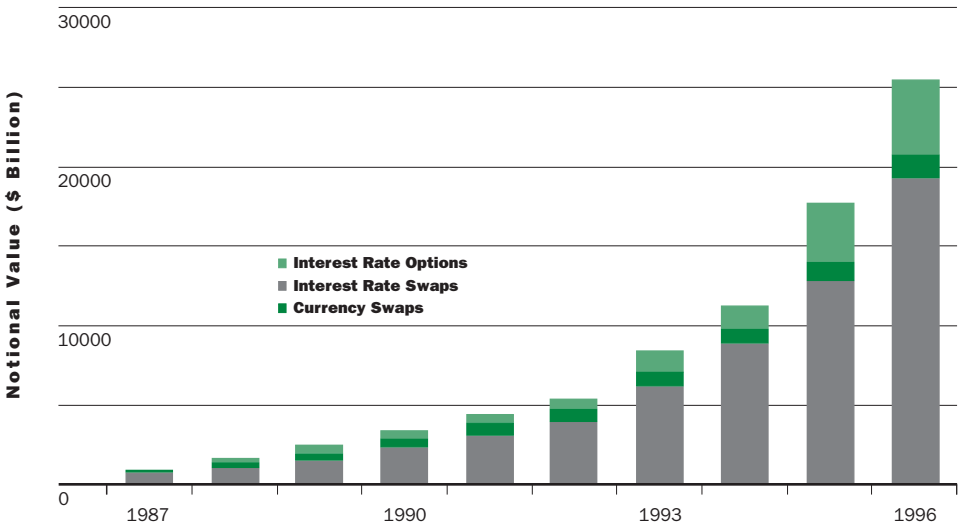
One of the reasons for such impressive growth rates in the volume of derivative transactions has been the ever-increasing demand for financial risk-management products by corporations. Although corporations from most industries report only very sketchy details of their risk-management strategies, there is a growing consensus that more and more firms are managing their exposure to various financial risks using derivative contracts. The top panel in the table reports the percentage of nonfinancial firms using derivatives contracts according to several recent empirical studies. For example, Dolde (1993) reports that 85 percent of

TABLE A
Users of Derivative Instruments

Author/Study	Firms Included/Surveyed	Percentage of Firms Reporting Derivatives Use
Nonfinancials		
1992 Mian	All nonfinancial firms with data on both LEXIS/NEXIS and Compustat. Number of firms: 3,022	25.5
1992 Dolde Survey	Survey of Fortune 500 companies Number of respondents: 244	85.0
1994 Wharton/Chase Survey	Survey of 2,000 nonfinancials not including Fortune 500 companies. Number of respondents: 530	35.0
1995 Wharton/Chase Survey	Survey of 2,500 nonfinancials including Fortune 500 companies. Number of respondents: 350	41.0
Banks		
Sinkey and Carter	All U.S. commercial banks, 1991 Number of banks: 11,308	5.4
	U.S. commercial banks with assets > $1 Billion, 1991 Number of banks: 353	75.9
Insurance Companies		
Cummins, Phillips, and Smith	All U.S life/health insurance companies, 1994 Number of life/health insurers: 1,202	9.8
	All U.S life/health insurance companies with assets > $1 Billion, 1994. Number of life/health insurers: 193	42.0
	All U.S property/casualty insurance companies, 1994 Number of property/casualty insurers: 1,664	6.7
	All U.S property/casualty insurance companies with assets > $1 Billion, 1994. Number of property/casualty insurers: 112	30.4

Sources: Bodnar, Hayt, and Marston (1996); Cummins, Phillips, and Smith (1997a); Dolde (1993); Mian (1996); and Sinkey and Carter (1994).

CHART A
Total Notional Value of Interest Rate Options, Interest Rate Swaps, and Currency Swaps Outstanding, 1987–96



Source: International Swaps and Derivatives Association, Inc.

the Fortune 500 companies responding to his survey use derivatives to manage their risk exposure.

There are a number of rationales for the increased demand for financial derivatives at the corporate level. First, it could be argued that it is less costly to write these contracts than it is to change the firm's operating or financial leverage. If this argument is true, the same features would make these instruments useful for managers who are prepared to take on additional risk with the hope of generating additional profits. For example, an insurance company may try to achieve a higher yield on its asset portfolio by investing in long-term, low-grade bonds. By purchasing such a security the insurer has an exposure to both movements in interest rates and movements in the credit quality of the borrowers. This net exposure can, however, be altered by purchasing interest rate derivatives, leaving the insurer with credit risk but not interest rate risk.

A second rationale for increased volume could involve the seminal work of Black, Scholes, and Merton on the pricing of options. These studies were published in the early seventies, about the same time that exchange-traded options were introduced in Chicago. Prior to this work, there did not exist a rigorous understanding of how to accurately price or use derivative securities. When combined with the fact that volatility in asset and commodity prices increased dramatically in the 1970s, '80s, and '90s (when compared with the earlier postwar years), one has all the ingredients needed to make these the popular financial instruments that they are today.

1. *The notional value of a derivative contract is analogous to the par, or face, value of an underlying contract as it is used to calculate the cash flows that change hands. It is not, however, necessarily the amount that is exchanged.*

results of various studies investigating the primary value-maximization rationales are presented below.

Mitigation of Financial Distress Costs. Numerous authors have investigated whether firms more likely to incur financial distress costs engage in risk management in an effort to reduce the probability of incurring these costs. The evidence is not persuasive for nonfinancial companies. An early study by Wall and Pringle (1989) reports that firms with lower credit ratings are more likely than higher-rated firms to use derivative contracts known as swaps.

Among the explanations that have been advanced to justify risk management as a value-maximizing decision is the need to mitigate the costs of financial distress, minimize taxes, and avoid costly external finance.

Other authors have considered the more general question of whether the firm's capital structure is related to the likelihood that the firm will engage in risk management via derivatives contracting. For example, neither Mian (1996) nor Nance, Smith, and Smithson (1993) report any evidence to suggest that derivatives trading is related to the capital structure of the firm. A more recent study by Geczy, Minton, and Schrand (1997) investigates the relationship between the capital structure of the firm and the decision to manage foreign currency exposures using derivatives. This study differs from its predecessors as the authors recognize the simultaneous nature by which managers make capital structure and risk-management decisions for their firms. Even after incorporating the joint decision-making process of managers in their estimation procedure, the authors conclude that there does not appear to be a relationship between the decision to use derivatives and capital structure choice.

One exception to these studies of nonfinancial firms is Dolde (1996). He finds that after controlling for the firm's underlying exposure to various financial risks, there is a significant complementary relationship between risk management and the leverage of the firm. That is, highly leveraged firms are more likely to use derivatives to avoid the expected costs of financial distress.

The evidence from studies investigating the decision by financial companies to use derivatives as a way to avoid financial distress costs is mixed. Sinkey and Carter (1994) provide only weak evidence suggesting that the capital structure and risk-management decisions of U.S. commercial banks are related. Likewise, Gunther and Siems (1995) report no significant relationship between the decision to use derivatives and the capital structure of the firm. In addition, focusing on only those banks that are active in derivatives markets, Gunther and Siems note that banks reporting a higher volume of derivatives activity also have higher capital ratios. This result is in fact inconsistent with the financial distress hypothesis, at least as it is usually defined in the literature. Cummins, Phillips, and Smith (1997b) find a similar result regarding the volume of derivatives activities for U.S. life/health insurers although they also report a significant and negative relationship between the capitalization level of both life/health and property/casualty insurers and the decision to use derivative securities, consistent with the financial distress hypothesis.

Use of Risk Management to Lower Expected Tax Burdens. Evidence on using risk management via derivatives contracting as a way to lower the firm's expected tax burdens is more convincing. Nance, Smith, and Smithson (1993) conducted one of the earliest empirical studies investigating whether taxes were a significant determinant of a firm's decision to transact in derivative markets. From their sample of nonfinancial companies, they conclude that firms with higher investment tax credits are more likely to engage in derivative transactions. Cummins, Phillips, and Smith (1997b) also find evidence consistent with the tax hypothesis. For the life insurance industry, they report a significant and positive relationship between the decision to participate in derivative markets and proxies for insurers having tax-loss carry forwards. They also find a positive relationship between derivatives usage and proxies for having net income in the progressive region of the tax schedule. Finally, a paper by Graham and Smith (1996) develops a simulation model to empirically determine the convexity of the tax schedule faced by a large sample of COMPUSTAT firms. They conclude that approximately 50 percent of the firms in their sample face convex tax schedules and therefore have an incentive to reduce the volatility of their income stream. They use the estimated simulation model and report that, for the subsample of companies that they estimate are facing convex tax functions, a 5 percent reduction in the volatility of the firm's taxable income stream leads to a 4.8 percent reduction in their expected tax liability.

Avoiding Costly External Financing. A number of authors have found strong evidence documenting that firms use derivatives to reduce the variability of their income stream and thus help ensure that adequate internal funds are available to take advantage of attractive projects. Gay and Nam (1997), for example, investigate nonfinancial companies' use of derivatives and provide test results consistent with the hypothesis that firms with both low levels of liquidity and high growth opportunities, as measured by the ratio of the market value to the replacement value of the firm, tend to hedge more. This finding is consistent with managers' trying to mitigate the need to seek costly external funds or lose their opportunity to invest in valuable projects.[12]

Other authors have found similar results. Studies of nonfinancial firms by Geczy, Minton, and Schrand (1997) and Nance, Smith, and Smithson (1993) both found that companies with less liquidity or companies that use less preferred stock, as opposed to using straight debt, are more likely to use derivatives to avoid circumstances under which a shock to the internal capital resources of the firm might force the company to forgo profitable projects.

A recent study by Ahmed, Beatty, and Takeda (1997) investigating 152 U.S. commercial banks also finds support for the costly external finance hypothesis. The authors report that banks with less liquidity are more likely to use derivatives to manage their exposure to various price risks. Finally, Cummins, Phillips, and Smith (1997a, 1997b) report that insurers with large proportions of their assets invested in illiquid markets, such as real estate for the property/casualty insurers or privately placed bond and collaterialized mortgage obligations for life insurers, are more likely to hedge the volatility of their income using derivatives.

Conclusion

This article has provided a review of the rationales that are often put forth concerning why corporations might engage in the practice of actively managing their exposure to a wide variety of risks—so-called risk-management practices. One school of thought is that managers attempt to reduce the volatility of cash flows because managers are personally averse to risk and their compensation is often tied to the firm's performance. Others have argued that managers attempt to overtly alter the risk profiles of their firms in an effort to increase the value of the firm's shares. However, basic finance theory says that, absent frictions in capital markets, shareholders can manage their own risk exposure. Thus, the value-maximization rationale for the use of derivatives requires some specific notion of important market imperfections because the use of insurance of this type is typically not free. Among the explanations that have been advanced to justify risk management as a value-maximizing decision is the need to mitigate the costs of financial distress, minimize taxes, and avoid costly external finance.

The discussion of the empirical literature on risk management focuses on one particular set of tools, namely, derivative securities. These contracts exist only for purposes of risk management and, as such, provide a natural set of data from which to glean managers' motives for changing the distribution of future cash flows. Tufano (1996) has provided some evidence from the gold mining industry that is consistent with the idea that managers use derivatives to reduce the volatility of their own income stream. Thus there is some evidence consistent with the managerial demand for risk management. On the other side of this question, the empirical evidence on the relationship between derivatives transactions and firm value has so far been mixed. However, there is a growing body of literature that suggests that at least a portion of total derivatives contracting is related to activities known to increase firms' value—for example, avoiding costly external finance and lowering expected tax bills. Further research on this question is important because it gets to the heart of whether or not derivatives in particular, and risk-management techniques in general, are being used to enhance value in underlying securities markets or to provide benefits to parties other than the shareholders of the firm.

12. The market-to-replacement value (or Tobin's Q) is a measure of growth opportunities used by a number of researchers. The logic is that if investors are willing to pay more than what it would cost to start the firm over, then they must believe that the firm's future prospects are valuable in an economic sense.

REFERENCES

AHMED, ANWER S., ANNE BEATTY, AND CAROLYN TAKEDA. 1997. "Evidence on Interest Rate Risk Management and Derivatives Usage by Commercial Banks." University of Rochester Working Paper, July.

BANK FOR INTERNATIONAL SETTLEMENTS. 1995. *Central Bank Survey of Derivatives Market Activity.* Basel.

BODNAR, G.M., G.S. HAYT, AND R.C. MARSTON. 1996. "1995 Wharton Survey of Derivative Usage by U.S. Non-Financial Firms." *Financial Management* 25:113–33.

BRENNAN, MICHAEL, AND EDWARDO SCHWARTZ. 1988. "The Case for Convertibles." *Journal of Applied Corporate Finance* 1:55–64.

COX, JOHN, AND MARK RUBINSTEIN. 1985. *Options Markets.* Englewood Cliffs, N.J.: Prentice Hall.

CUMMINS, J. DAVID, RICHARD D. PHILLIPS, AND STEPHEN D. SMITH. 1997a. "Corporate Hedging in the Insurance Industry: The Use of Financial Derivatives by U.S. Insurers." *North American Actuarial Journal* 1:13–49.

———. 1997b. "Derivatives and Corporate Risk Management: Participation and Volume Decisions in the Insurance Industry." Paper presented at the 1997 conference of the International Association of Financial Engineers, Boston, September 22–24, 1997.

DOLDE, WALTER. 1993. "The Trajectory of Corporate Financial Risk Management." *Journal of Applied Corporate Finance* 6:33–41.

———. 1996. "Hedging, Leverage, and Primitive Risk." *Journal of Financial Engineering* 4:187–216.

FROOT, KENNETH A., DAVID S. SCHARFSTEIN, AND JEREMY C. STEIN. 1993. "Risk Management: Coordinating Invest-ment and Financing Policies." *Journal of Finance* 68:1629–58.

GAY, GERALD D., AND JOUAHN NAM. 1997. "The Under-investment Problem and Derivatives Usage by Corpora-tions." Paper presented at the 1997 conference of the International Association of Financial Engineers, Boston, September 22–24, 1997.

GECZY, CHRISTOPHER, BERNADETTE A. MINTON, AND CATHERINE SCHRAND. 1997. "Why Firms Use Currency Derivatives." *Journal of Finance* 52:1323–54.

GRAHAM, JOHN R., AND CLIFFORD W. SMITH. 1996. "Tax Incentives to Hedge." University of Rochester Working Paper, September.

GROSSMAN, SANFORD J., AND OLIVER D. HART. 1981. "The Allocational Role of Takeover Bids in Situations of Asymmetric Information." *Journal of Finance* 36:253–70.

GUNTHER, JEFFERY W., AND THOMAS F. SIEMS. 1995. "The Likelihood and Extent of Bank Participation in Derivative Activities." Federal Reserve Bank of Dallas Working Paper, May.

MACMINN, RICHARD D. 1987. "Forward Markets, Stock Markets, and the Theory of the Firm." *Journal of Finance* 42:1167–85.

MIAN, SHEHZAD L. 1996. "Evidence on Corporate Hedging Policy." *Journal of Financial and Quantitative Analysis* 31:419–39.

MILLER, MERTON H., AND FRANCO MODIGLIANI. 1961. "Dividend Policy, Growth, and the Valuation of Shares." *Journal of Business* 34:411–33.

MILLER, MERTON H., AND DAVID J. ROSS. 1997. "The Orange County Bankruptcy and Its Aftermath: Some New Evidence." *Journal of Derivatives* 4:51–60.

MODIGLIANI, FRANCO, AND MERTON H. MILLER. 1958. "The Cost of Capital, Corporation Finance, and the Theory of Investment." *American Economic Review* 48:261–97.

NANCE, DEANA R., CLIFFORD W. SMITH JR., AND CHARLES W. SMITHSON. 1993. "On the Determinants of Corporate Hedging." *Journal of Finance* 68:267–84.

OVERDAHL, JAMES, AND BARRY SCHACHTER. 1995. "Derivatives Regulation and Financial Management: Lessons from Gibson Greetings." *Financial Management* 24:68–78.

PETERSON, MITCHELL A., AND S. RAMU THIAGARAJAN. 1997. "Risk Measurement and Hedging." Northwestern University, J.L. Kellogg Graduate School of Management,Working Paper.

SHARPE, WILLIAM. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance* 19:425–42.

SINKEY, JOSEPH F., AND DAVID CARTER. 1994. "The Determinants of Hedging and Derivatives Activities by U.S. Commercial Banks." University of Georgia Working Paper presented at the American Finance Association Annual Meeting, Washington, D.C., January 6, 1995.

SMITH, CLIFFORD W., JR., AND RENE M. STULZ. 1985. "The Determinants of Firms' Hedging Policies." *Journal of Financial and Quantitative Analysis* 20:391–405.

SMITH, DONALD J. 1997. "Aggressive Corporate Finance: A Close Look at the Proctor & Gamble/Bankers Trust Leveraged Swap." *Journal of Derivatives* 4:67–79.

TUFANO, PETER. 1996. "Who Manages Risk? An Empirical Examination of Risk Management Practices in the Gold Mining Industry." *Journal of Finance* 51:1097–137.

U.S. GENERAL ACCOUNTING OFFICE. 1997. "Financial Derivatives: Actions Taken Since May 1994." Washington, D.C.

WALL, LARRY D., AND JOHN PRINGLE. 1989. "Alternative Explanations of Interest Rate Swaps: An Empirical Analysis." *Financial Management* 18:119–49.

The Impact of Fraud on New Methods of Retail Payment

WILLIAM ROBERDS
The author is a research officer in the macropolicy section of the Atlanta Fed's research department.

IN MARKET ECONOMIES, PAYMENTS SYSTEMS PROVIDE CERTAINTY OF VALUATION IN EXCHANGE. PEOPLE SELLING GOODS OR SERVICES EXPECT MONEY IN RETURN, WHERE *MONEY* MEANS EITHER CURRENCY OR A FINANCIAL CLAIM THAT IS WORTH A FIXED AMOUNT OF CURRENCY. TO PROVIDE THIS CERTAINTY, A SUCCESSFUL PAYMENTS MEDIUM HAS TO OVERCOME VARIOUS RISKS THAT ARE A NATURAL PART OF THE PAYMENTS PROCESS.

An important risk associated with payments systems is the risk of fraud. Fraud can occur because purchases of goods typically involve at least three parties. The first party, a buyer (sometimes referred to as a consumer), wants to purchase some good or service from the second party, a seller (or merchant). In modern economies, such purchases are rarely accomplished by barter, or direct trade of goods between buyers and sellers. Instead, the buyer offers to transfer to the seller a claim on a third party, an issuer.

Such transactions are preferable to barter because it is easier for sellers to value such claims than to value goods offered in barter. If, however, the issuer cannot be physically present to verify the claim when it is transferred from buyer to seller, then there is always some chance that the buyer may offer a fraudulent claim.[1]

Payments fraud takes on many forms, but most cases of fraud consist of one of two types of misrepresentation. The first is an offer to exchange a claim where none exists. For example, a buyer may write a check on insufficient funds. The second type of misrepresentation occurs when a buyer offers to transfer a claim that rightfully belongs to someone else. Examples of this type of fraud include check forgery or use of a stolen credit card.

As these examples indicate, traditional payments media such as currency, checks, and credit cards are not exempt from the risk of fraud. Currency fraud (counterfeiting), check fraud, and credit card fraud are serious problems, costing the U.S. economy billions of dollars each year. But with each of these payments methods, the problem of fraud has been kept at a manageable level so that their overall integrity has been maintained.

This article explores the potential impact of fraud on new forms of retail payment such as electronic cash and stored-value cards. These new payments media can increase economic efficiency by incorporating advances in computer technology into payments systems. Payments systems based on these new media communicate much of the same information as traditional payments systems but at a potentially lower cost. Electronic payments systems have this advantage because it is cheaper to move electrons than it is to move paper. This natural advantage of electronic systems can be a disadvantage, however, when it comes to the risk of fraud. Since computer data are readily stored, copied, and manipulated, complex security procedures are needed to guarantee the integrity of electronic payments data.

Will the risk of fraud hinder the development of the new payments media? This article investigates this issue by first considering which features of payments media are conducive to fraud. The discussion then turns to which of these features are also part of traditional payments systems. Finally, the article considers some new payments media, how certain features of these media differ from more traditional forms of payments, and whether these features are likely to detract from the acceptance of the new media in the marketplace.

The Optimal Incidence of Fraud

Any discussion of payments fraud should begin from the basic economic principle of balancing costs and benefits. That is, the benefits of measures designed to reduce fraud should exceed the costs of such measures.

It is technologically possible to virtually eliminate fraud in electronic payments, as has been demonstrated by the experience of large-value, or wholesale, funds transfer systems. Such systems, which are used by banks and securities markets participants, are practically devoid of fraud. However, large-value systems typically make use of elaborate and costly security measures (for example, using dedicated telephone lines for all transactions) that would be excessively costly, time-consuming, or otherwise inappropriate for retail payments systems.

Thus, the key question for retail payments systems is not whether fraud will occur but instead how much fraud can be tolerated if the payments system is to remain effective. While this amount will most certainly be positive, both economic intuition and practical experience suggest that the optimal amount of fraud is relatively small. Intuitively, fraud is particularly injurious to the provision of payments services because it detracts from the essential quality of the service that is being provided, which is certainty of valuation in exchange. This intuition is backed by the experience with traditional payments media, for which fraud rates are far from negligible but, nonetheless, relatively low.

Incentives for Fraud

The problem of fraud is common to all payments systems and dates back to ancient times. Nonetheless, there are some types of transactions and some features of payments systems that are more likely than others to create incentives for fraud. Some of the key factors influencing the risk of fraud are the following.

Face Value of the Claim. For fraud to be profitable, the reward from committing fraud has to be large enough to offset the threat of punishments imposed by the legal system. There is little incentive to create fraudulent small-denomination claims such as coins. On the other hand, transactions of sufficiently large value are more likely to inspire the use of costly security measures, as noted above.

Verifiability. If the existence and ownership of a claim can be instantly verified, say, through an on-line verification system, this ability obviously reduces the risk of fraud. Effective verification systems are costly to set up and operate, however.

Payments systems based on these new media communicate much of the same information as traditional payments systems but at a potentially lower cost.

Anonymity of the Transaction. If a buyer and seller do not have an ongoing business relationship, the incentive for fraud increases. The incentive for fraud is also enhanced if the ownership of the claim offered in payment cannot be traced.

Point-of-Sale Transactions. If the seller can withhold delivery of the good until the claim can be verified, then the incentive for fraud is reduced. If the good is exchanged at the point of sale, there is always some chance that the claim presented by the buyer is fraudulent.

Allocation of Losses. Perhaps the most critical factor contributing to the incidence of fraud is the allocation of losses.

Suppose that a fraudulent transaction has occurred. Who should bear the costs of the fraud? Note that to the extent that prices must be raised in order to cover losses from fraud, all market participants may end up bearing some of the cost. However, having different rules concerning the allocation of loss from a particular incident of fraud changes the distribution of losses among individual buyers, sellers, and issuers and hence affects the incentives to commit fraud.

One possibility is that these costs are borne directly by an individual buyer. This arrangement gives maximum reassurance to the seller and to the issuer. In some cases, however, the buyer and the legitimate owner of the transferred claim may be two different people. For example, in

1. *A second possibility is that the seller could offer worthless merchandise, which is a potentially serious problem with some forms of electronic commerce. Yet another possibility is that issuers could issue claims on worthless assets. New forms of financial intermediation are not immune to this type of risk, as evidenced by the recent collapse of the European Union Bank, an "Internet bank" based in Antigua (see Rohter 1997). Nonetheless, this article will focus on the first risk as the most likely to affect acceptance of new payments media.*

the case of a check forgery, the forger does not have ownership of the claim (deposit) apparently represented by the forged check. And in point-of-sale transactions, the buyer may be long gone by the time that a fraud is discovered. For these reasons it can be problematic to assign the costs of fraud to buyers.

A second possibility is for the seller to bear the costs of fraud. This arrangement protects the interests of the buyer and the issuer, but it is unlikely to be popular with sellers.

The third possibility is that the issuer of the claim bears the costs. This is clearly the most convenient arrangement for the buyer and seller, but it is also the most likely to promote fraud. Since the issuer is not present at the transaction, the legitimacy of a claim on the issuer can never be verified with absolute certainty.

In spite of this disadvantage, there are many circumstances in which it makes sense for the issuer to bear the risk associated with fraud. If the wealth of the issuer is large, compared with that of the buyer (say, a typical consumer) and the seller (say, a small business), then the issuer may be the party most prepared to face such risks. A large issuer may also be able to lessen exposure to fraud risk by diversifying this risk over many transactions.

The above discussion suggests that the problem of fraud will be greatest in cases involving large informational asymmetries between buyer and seller. Fraud is more likely to occur when transactions involve large amounts, when verification is costly, in anonymous transactions, and in point-of-sale transactions. Fraud will also be more likely in transactions in which at least some of the costs of fraud can be shifted to the third party, the issuer of the claim used for purchase.

Fraud and Traditional Payments Systems

Currency. The simplest traditional payments system is currency. In modern-day currency transactions, the role of issuer is played by a central bank or sovereign government.[2] The claim in this case is a fixed-denomination note or coin that is considered a liability of the issuer. Payment is effected by physical transfer of the note or coin. Among traditional payments systems, currency is unique in that a payment in currency does not need to be cleared and settled through the banking system in order to constitute a valid payment. Another distinguishing feature of currency is that it can circulate indefinitely before it is returned to its issuer.

Fraud can occur in currency transactions if the currency is counterfeit or stolen. The fact that currency is a convenient, widely accepted, and anonymous medium for point-of-sale transactions in turn creates incentives for counterfeiting and theft.

Several factors serve to limit the risk from counterfeiting currency, however, at least within the United States. The first is vigorous law enforcement; according to the U.S. General Accounting Office (GAO) (1996), the majority of counterfeit currency is seized before it can be distributed. The second factor is that since all detected counterfeit currency is subject to seizure by law enforcement authorities, a significant portion of the costs of counterfeit fraud is borne by buyers and sellers. The third factor is that currency is not widely used within the United States for transactions with a high dollar value because other, more suitable payments systems are widely available. Anyone attempting to pass a large amount of counterfeit currency would be forced to use it in a large number of small-value transactions.

The problem of theft also tends to be self-limiting. Since currency is anonymous, a buyer holding a large amount of cash is liable for its theft or loss. Consequently, most people do not hold large amounts of currency.

Statistics on the incidence of counterfeiting are difficult to obtain since counterfeit currency can circulate for some time without being detected. Available statistics suggest that counterfeiting is not an economically significant problem in the United States. In 1994 the total amount of counterfeit currency detected by law enforcement was less than one-tenth of 1 percent of currency outstanding, most of which never reached circulation (GAO 1996, 11).

Checks. Payment by check is by far the most prevalent system for noncurrency retail payments in the United States. In a check transaction, a buyer instructs a bank or similar financial institution to transfer the buyer's deposit claim on a bank. The buyer does so by transferring an order to pay, or check, to the seller. The seller or seller's bank then presents the check to the buyer's bank for payment.[3] In such a transaction, the bank plays the role of issuer, although the check is considered a liability of the buyer and not of the bank on which it is drawn.[4]

Checks are a natural target for fraud as they can be written for large amounts, are relatively easy to alter or forge, and can be difficult or costly to verify at the point of sale. Check fraud has recently become a more serious problem because of several factors. The first is the widespread availability of computer technology, which has made it easier to counterfeit checks (see, for example, Hansell 1994 or Nielsen 1994). The second factor has been the funds availability schedules required by the Expedited Funds Availability Act of 1987 (see Board of Governors 1996b). The act requires that banks make check funds available according to certain, preset schedules. Consequently, banks must sometimes make funds available before they can ascertain whether a deposited check is fraudulent.

Despite these problems, there are certain factors that have served to limit the incidence of check fraud. The first and most important is the allocation of losses. While the law governing the allocation of losses from

check fraud is complex, the end result is that the liability for fraud often resides with the seller and not the bank on which the check is drawn.[5] For example, a merchant who accepts a check in a point-of-sale transaction bears the loss if the check is returned for insufficient funds.[6] Likewise, if a check is stolen, a buyer can stop payment on the check, again leading to potential losses for the seller. As a result of this loss allocation, there is widespread recognition of the potential for fraud in check transactions and sellers are reluctant to accept checks in situations that are conducive to fraud, such as anonymous, point-of-sale transactions.

A second factor limiting the incidence of check fraud has been the increased use of techniques such as positive pay. Under a positive pay arrangement, a buyer (typically a corporation) sends a list of issued checks to the buyer's bank. Only checks on the list are automatically paid by the bank. Any check not on the list requires explicit approval by the buyer before it can be paid. Positive pay has been an effective weapon against losses resulting from check counterfeiting, forgery, and embezzlement, among others. A third factor has been the Federal Reserve's requirement for "large-dollar return notifications." That is, banks must provide prompt notice of nonpayment on checks for $2,500 or more. Prompt notice of nonpayment reduces the likelihood that banks will provide provisional credit for fraudulent checks before the fraud can be discovered.

As is the case with currency, available statistics suggest that check fraud is not a large enough problem to significantly detract from the use of checks as a payments medium. Estimates of the total cost of check fraud in the United States range as high as $10 billion annually (Hansell 1994). An extensive 1995 survey by the Federal Reserve found that banks' share of these losses amounted to $615 million in 1995.[7] While these figures show that check fraud is a serious problem, these numbers are small compared with the total volume of check payments in the United States, which was roughly $73.5 trillion for 1995 (Bank for International Settlements 1996b). The overall rate of check fraud loss is less than 2 basis points, or two-hundredths of 1 percent.[8]

Credit Cards. Credit cards are widely used in retail payment situations, especially when informational asymmetries make payment by check impractical. In a credit card transaction, the buyer pays for a purchase by drawing on a line of credit from the credit card issuer. The issuer pays the seller for the purchase, and the balance on the credit card is then paid down by the buyer. Since the claim presented in payment is considered a liability of the credit card issuer, this type of transaction transfers much of the risk of insufficient funds in the original transaction from the seller to the credit card issuer.

In cases of credit card theft or similar types of fraud, cardholders' liability is restricted by the Truth in Lending Act of 1968 and corresponding Federal Reserve Regulation Z. Generally a cardholder's liability is limited to $50 as long as the cardholder reports a lost or stolen card, and in practice the liability is often less than this maximum. The remaining liability is shared between the seller, or merchant, and the credit-card issuer. While the rules governing the apportionment of this liability vary, the GAO (1997, 114) reports that, on average, the vast majority (70 percent) of the liability is borne by the credit card issuers. To limit incentives for fraud, the issuer's liability is contingent on the merchant taking certain steps intended to curtail fraud (for example, validating a credit card transaction through an on-line verification system).

The key question for retail payments systems is not whether fraud will occur but instead how much fraud can be tolerated if the payments system is to remain effective.

The incidence of fraud in credit card purchases is quite small in absolute terms but is relatively high as compared with checks. While precise figures are unavailable for the credit card industry as a whole, one estimate put total (gross) fraud losses at $2 billion to $3 billion in 1993 (Pearsall 1994), and another placed this figure at $1.3 billion for 1995 (Fryer 1996). Given aggregate credit card use of $879 billion for 1995, the estimates imply a fraud rate of between 10 and 20 basis points (0.1 to 0.2 percent). In the case of bank cards (MasterCard and Visa), a study by the American Bankers Association (1996) estimated total gross fraud loss for 1995 at $790 million versus purchases of $451 billion, implying a loss rate of 18 basis points (0.18 percent).

2. Historically such notes were also issued by commercial banks. These notes are discussed on page 48.
3. For an introduction to details of check clearing and settlement, see GAO (1997).
4. Exceptions are traveler's checks, cashier's checks, and certified checks.
5. Generally the loss allocation is determined by Articles 3 and 4 of the Uniform Commercial Code.
6. Of course, in such cases the merchant is entitled to try to recover the amount of the check through legal action.
7. See Board of Governors (1996b, 5). A smaller survey by the American Bankers Association (1994) put this number at $815 million for 1993. Both numbers represent "gross losses," that is, they do not incorporate any recoveries of lost funds.
8. This is an average rate for all checks, many of which are at low risk for fraud. The risk of fraud is substantially higher for certain types of checks.

Note that the relatively high rate of fraud on credit cards does not reflect any inherent shortcoming of credit cards as a payments medium. Rather, the fraud rate on credit cards reflects the fact that credit cards tend to be used in situations where incentives for fraud are greater, particularly in point-of-sale transactions. The acceptance of credit cards in such situations, together with the fact that the card issuers bear the majority of costs associated with fraud, help make credit cards a secure and convenient payments medium from the standpoint of marketplace participants.

To limit the potential for fraud, credit card issuers have invested heavily in on-line verification technology and other technologies to detect fraudulent use (see Fryer 1996 or Rutledge 1996). While this technology has been effective, it is also costly: Caskey and Sellon (1994) report that credit cards are the most expensive medium for retail transactions.[9]

Debit Cards. Conceptually, a debit card transaction closely resembles a check transaction. In a debit card transaction, a buyer transfers deposit claims from the buyer's bank account to that of the seller, just as in a check transaction. As with checks, this transfer is done as a *debit* transaction, in which funds are "pulled" by the seller (via the card network) from a buyer's bank account.

However, there are several key differences between a debit card transaction and a check transaction. The most important is that in contrast to most check transactions, the transaction itself is subject to an electronic verification process, which varies according to the type of card.[10] This verification process lessens the credit risk associated with the transaction. A second key difference is that a debit transaction is cleared and settled electronically through the card issuer's network rather than through a traditional paper-based check-clearing process. That is, in contrast to checks, the clearing and settlement of transactions does not have to wait for physical delivery or *presentment* of checks but can begin more or less immediately.

Debit card transactions also differ from credit card transactions in that the amount of a purchase is automatically debited from the buyer's bank account within a few days of the time of purchase. By contrast, credit card holders have to either pay for purchases after a grace period or pay interest on the unpaid balance.

In cases of debit card fraud, cardholders' liability is limited by the Electronic Funds Transfer Act of 1978 and the corresponding Federal Reserve Regulation E. Losses are capped at $50 if loss or theft of a debit card is reported within two days and at $500 if the loss is reported within sixty days. Recently the two main debit card issuers, MasterCard and Visa, have announced policies that place more stringent limits on cardholders' liability (see Fickenscher 1997 and Keenan 1997). Under these new policies, cardholders' liability is generally limited to $50. Available estimates suggest that the overall rate of fraud for debit card purchases is quite low, comparable to that for credit card purchases (Lunt 1996 and Keenan 1997).

Why Things Might Be Different with New Payment Technologies

Recently a number of new retail payment technologies have become available (some of which are still undergoing trial). Among the most widely discussed technologies are stored-value cards and a group of technologies that fall under the term *on-line payments.*[11]

A stored-value card is a payment card similar in appearance to a credit or debit card. To use a stored-value card, a buyer must first purchase a card from an issuer. The issuer then stores the value of this purchase on the card itself, in the form of data contained on a magnetic stripe or an electronic chip. A buyer can then purchase goods by presenting the card to a seller, who electronically transfers the value on the card to the seller's card or account. The value on the card must eventually be redeemed by the issuer.

On-line payments technology includes a number of important payments media, including on-line banking, on-line credit card payments, and electronic cash. On-line banking allows consumers direct computer access to banking services, either through "closed" networks such as traditional Automated Teller Machine networks or, more recently, through "open" networks such as the Internet. Using on-line banking, a buyer can initiate payment in much the same way as by writing a check. Clearing and settlement of on-line payment instructions often takes place via the automated clearinghouse system (the electronic interbank payments system for small-value transactions). In on-line credit card payments, a buyer initiates a credit card transaction by sending the buyer's credit card information to a seller over a computer network (almost always the Internet). Finally, payments can be made over the Internet by transfer of electronic cash, a difficult-to-counterfeit series of electronic messages that represent a financial claim on its issuer.[12]

In many ways, these new forms of payment closely resemble traditional forms. For example, stored-value cards have many features in common with travelers' checks, and credit card payments over the Internet are obviously not so different from credit card payments made at the point of sale or over the telephone. There are some features of the new payments media, however, that are not incorporated into traditional modes of payment. Some of these may affect the incidence of fraud and are discussed below.

One noteworthy feature of many of the new payments media (on-line credit card payments, some forms of on-line banking, and electronic cash) is that they allow for payments over the Internet, which is an open system

of computer networks with few restrictions to access.[13] The key advantage of the Internet over closed systems is that it allows buyers and sellers low-cost access to a greater range of transactions. While its open architecture makes the Internet an appealing vehicle for electronic commerce, this same openness offers opportunities for counterfeiting and fraud. The fact that a buyer or seller is on the Internet proves nothing in and of itself; additional verification of the transaction is required. For example, buyers making credit card purchases over the Internet need to convey enough information to show that credit cards offered in payment are not counterfeit or stolen. At the same time, sellers need to demonstrate that they are selling a legitimate product and not just collecting credit card numbers for fraudulent use. And both buyers and sellers need to safeguard against surreptitious monitoring of transactions by third parties. The need to verify on-line transactions has led to the development of technologies such as the Secure Electronic Transactions (SET) protocol (see, for example, Bloom 1997 or "Survey of Electronic Commerce" 1997). These technologies are designed to allow buyers and sellers to identify one another over the Internet and also to prevent unwanted eavesdropping on private transactions.

A similar difficulty exists with stored-value cards. These cards are designed to be used for small-dollar-value transactions, particularly transactions in which traditional methods of verification are too costly or otherwise impractical. Instead, verification is provided by data contained on the card itself (perhaps in combination with on-line information). In this sense, stored-value payments systems may be seen as an electronic analog of currency, where validity of the payments medium is provided by visual inspection. As is the case with currency, this feature of stored-value cards increases the incentives for counterfeiting and fraud. Stored-value systems rely on electronic encryption technologies to protect against counterfeiting and other fraudulent use.[14]

Incentives for fraud are magnified in the case of those stored-value cards that allow for "peer-to-peer" transactions, that is, transactions among cardholders who do not have access to on-line verification or clearing technologies. In this type of system, value can be successively transferred from one stored-value card to another without outside verification. This feature can increase the time interval between counterfeiting or possible fraudulent use of the card and the subsequent detection of fraud when the stored value is ultimately presented for redemption.

The issue of who bears the responsibility for fraud is unresolved for many of the new payments media. For many of these media, however, there are strong justifications for the issuer bearing the responsibility for losses due to fraud. The presence of "network effects" in payments technologies means that new forms of payment are unlikely to be issued by a single financial institution but instead by consortiums of financial institutions, data processing firms, and so on, operating under a single "brand name."[15] A network effect occurs when the entrance of one participant into a payments network increases the benefits or lowers the costs of participating in the network for all other network members. For example, if only one merchant in a small town accepts a particular brand of stored-value card, then consumers might not find it advantageous to use this card, making it difficult for the card issuer to recover costs. If, on the other hand, all the merchants in the same town were to accept this card, then consumers would be more likely to use the card regularly, which would in turn increase its profitability. Since the usefulness and profitability of a branded payments network depends heavily on its widespread acceptance, "branded" networks have a natural incentive to absorb the risk associated with fraud losses.

The general feeling expressed by policymakers is that the long-run benefits to the development of new payments technologies will outweigh any short-term difficulties associated with their introduction.

9. Fraud represents a significant, though relatively minor, component of this cost differential. A more significant component is the cost of delinquencies (failure to pay accounts due). Delinquencies in 1995 amounted to 3.55 percent of outstanding credit card balances, according to the American Bankers Association (1996).

10. Debit cards may be either "on-line" or "off-line." With on-line cards, a transaction is verified by comparing the purchase amount against a buyer's bank balance. With off-line cards, the transaction is verified by comparing the buyer's total purchases over a certain period against a preset limit.

11. These technologies are extensively discussed in Congressional Budget Office (1996), U.S. Department of the Treasury (1996), and GAO (1997).

12. Electronic cash is also known as e-cash, digital cash, electronic scrip, and electronic coins.

13. See McAndrews (1997a) for an introduction to the Internet and its potential uses in electronic commerce.

14. Encryption refers to the use of mathematical algorithms to convert data into a coded form. See Bank for International Settlements (1996a) on the use of encryption in payments systems.

15. A detailed discussion of this scenario is laid out in McAndrews (1997b). More generally, see Weinberg (1997) on network effects in payments systems.

Counteracting this incentive are potential difficulties resulting from anonymity of some of the new payments media, particularly for some stored-value cards. For example, if a stored-value card is issued anonymously there is no way to identify the rightful owner of the card. It would thus be difficult if not impossible for the issuer of the card to stop payment on a lost or stolen card, given the current design of stored-value systems.[16] In such situations, users of stored value cards would have an incentive to handle these cards with the same care as if they were currency.

What Could Go Wrong?

A recent episode in Japan provides some sobering lessons concerning the potential for fraud over new payments systems (see Glain and Shirouzu 1996 and Pollack 1996). This case concerns a stored-value card designed by Sumitomo Corporation and Mitsubishi Corporation, with the cooperation of Nippon Telephone and Telegraph as well as various government agencies. The cards were intended for use with pachinko, a type of pinball game. One purpose of the cards was to limit criminal activities often associated with the pachinko parlors, such as gambling, tax evasion, and money laundering.

The value on the cards was held in the form of data stored on magnetic strips.[17] Criminal organizations were able to defeat the encryption by *cloning,* that is, by transferring the data stored on existing cards to used cards. The cloned stored-value cards were then taken to pachinko parlors and redeemed for cash. Since the stored-value issuers had no way to distinguish fraudulent transactions from legitimate transactions, they were forced to absorb the resulting losses. Published reports estimate the losses from this episode were at least $600 million.

The pachinko fraud is instructive in that it illustrates the power of incentives. Although the pachinko stored-value cards were heavily encrypted, various features of their design created strong incentives for fraud. Apart from the obvious defect of being too easy to copy, the cards were almost perfectly anonymous, were designed for point-of-sale transactions, and were available in large denominations (of about $50 and $100). Pollack (1996) reports that reductions in fraud were achieved only after the card issuers both improved the cards' encryption technology and reduced the incentives for fraud by eliminating large-denomination cards and cracking down on pachinko parlor operators who had apparently tolerated extensive use of cloned cards.[18]

Historical Lessons

Various analyses of new payments media (particularly stored-value cards and electronic cash) have invoked comparisons of the new media with the banknotes that circulated during the U.S. Free Banking Era (1837–65).[19] During this period, banks issued claims in the form of bearer notes, which circulated much as government-issued currency does today. Banknotes usually traded at par value locally but were often traded at a discount in transactions that occurred at any distance from the issuing bank.[20] A major cause for this discounting was the fraud risk associated with counterfeit and altered notes.[21] Given that certain of the new electronic payments media share a number of features with privately issued banknotes, would we expect a similar pattern of discounting to arise? The most likely answer to this question is no, for at least two reasons.

First, Free Banking Era banknotes were particularly attractive targets for fraud. Often the notes were available only in large denominations ($5 and up, the equivalent of roughly $80 today), they were widely used for anonymous, point-of-sale transactions, and nonlocal notes could only be verified at considerable cost and after a lengthy delay.[22] This unfortunate combination of features is not shared by any of the new payments media.

Second, Gorton (1996) shows that despite the prevalence of fraud, the most serious risk to holders of Free Banking Era banknotes, and hence the greatest source of discounting, was not fraud risk but credit risk associated with the issuer. In this case, credit risk refers to the risk that a note would not be honored at full value because of either the insolvency or illiquidity of the issuing institution. During the Free Banking Era, banknotes' credit risk was exacerbated by a combination of poor communications and restrictive banking laws. These laws effectively prohibited banks from branching beyond their home state or local area, thereby making it difficult for banks to build effective coalitions in order to guarantee the value of their notes. In New England, where banks were able to form such a regional coalition, discounting of notes on banks within the coalition was practically nonexistent.[23] The experience of the New England banks suggests that if the credit risk associated with a payments instrument can be held in check, then fraud risk is unlikely to lead to discounting of that instrument.

As discussed above, the "network" economics of the new payments media are likely to limit credit risk associated with new forms of payment. Holders of stored-value cards, for example, would prefer to use stored-value cards that are readily acceptable in as many places as possible. Providers of stored-value cards and similar payments systems therefore have incentives to form broad coalitions with a widely recognizable brand name. The members of such coalitions have strong incentives to monitor each others' credit risk in order to maintain credibility of the brand.

Credit risk could also be eliminated by Federal Deposit Insurance Corporation (FDIC) insurance of a payment instrument. As of this writing, however, it appears that FDIC insurance will not be provided for most types of stored-value cards. The FDIC has also requested

comment on the eligibility of certain other forms of electronic payment for deposit insurance; see FDIC (1996).

The Free Banking Era experience suggests that a necessary downside of containing credit risk may be increased fraud risk, however. According to Gorton (1996, 370) the banknotes of established, creditworthy banks were the most likely targets of counterfeiters. Notes of less creditworthy banks were more likely to be discounted and less likely to circulate, and hence they were not worth the trouble.

Public Policy Concerns

An important challenge for policy in the area of new payments technologies has been to promote increases in efficiency associated with technological improvements while safeguarding consumers from undue risks. To date, public policy toward new forms of retail payment has been largely hands-off. The general feeling expressed by policymakers is that the long-run benefits to the development of new payments technologies will outweigh any short-term difficulties associated with their introduction. The view has also been expressed that premature regulation of new payments media may hinder the development of potentially more efficient payments systems.[24]

In the case of stored-value cards, the Federal Reserve has attempted to avoid excessive regulatory burdens on new payments technology by proposing that Regulation E not apply to certain types of stored-value cards (see Board of Governors 1996a, 1997a). Currently, Regulation E requires that consumers be provided written records for electronic funds transfers and limits consumer liability to $50 (see discussion above) when they use an "access device" to withdraw or transfer funds from a "consumer asset account." While withdrawals from such an account in order to load the stored-value card would be covered by Regulation E, the proposed regulations would not be extended fully to all transactions between buyers and sellers involving stored-value cards. For example, the Federal Reserve proposal would exempt from these provisions all cards containing $100 or less, as well as all cards that are off-line and do not track individual transactions.

Another important public policy issue in this area has to do with potential trade-offs between security and privacy. As discussed above, one of the factors affecting the risk of payment is the anonymity of the transaction. If a seller has access to enough information about a potential buyer (for example, the buyer's current bank balance), then the risk of fraud can be minimized. On the other hand, a seller's need for information about the creditworthiness of potential customers can conflict with the customers' need for privacy.

Payments systems that make use of extensive consumer-identifying information can lessen the incidence of fraud . . . but the value of such information in reducing fraud must be balanced against the value of privacy.

This conflict of interest has become more acute in recent years. Improvements in computing and communications technology have enabled the construction of extensive computer databases of information on consumers.[25] Widespread use of electronic payments media could result in the creation of even more extensive databases, providing detailed information on the purchasing habits of users of new payments media. While there would be many legitimate uses of such information, including abatement of fraud risk, its use could also result in some loss of privacy.

In some cases, identifying information on consumers has served to enable, rather than to deter, fraud.

16. *See Task Force on Stored-Value Cards (1997, 715–20) or Board of Governors (1997a, 52) for a discussion of these issues.*
17. *Stored-value cards that make use of data stored on magnetic strips are generally viewed as less secure than cards on which the data is stored on an electronic chip.*
18. *While it was difficult to detect individual fraudulent cards during this episode, the widespread use of such cards was public knowledge. According to Pollack (1996), the scale of the fraud became evident when long lines of people would form outside of certain pachinko parlors, hours before the parlors were open for business.*
19. *See, for example, Greenspan (1996), Dwyer (1996), Rolnick, Smith, and Weber (1997), McAndrews (1997b), or Schreft (1997).*
20. *Merchants used publications known as "banknote reporters" to keep track of the notes' current market value.*
21. *See, for example, Dillistin (1949) or Gorton (1996) on the prevalence of note fraud during the Free Banking Era.*
22. *At the time, restriction of note issue to large denominations was thought necessary to lessen the incidence of note fraud; see White (1995) for a discussion. The reasoning was that holders of small notes would lack sufficient incentive to check on their authenticity. Another motive for restricting issue of small-denomination notes was the fear that their issue would lead to inflation and ultimately to erosion of the gold standard; see Timberlake (1978, chap. 9). See Sargent and Wallace (1982) for a modern interpretation of this view.*
23. *The regional coalition of New England banks was known as the Suffolk System. See, for example, Calomiris and Kahn (1996) or Rolnick, Smith, and Weber (1997) on the operation of the Suffolk System.*
24. *See, for example, Blinder (1995), Kelley (1996), Greenspan (1996), and Kamihachi (1997).*
25. *See Board of Governors (1997b) or Bernstein (1997) for examples of commercially available data on consumers.*

In these "identity theft" cases, criminals have been able to use stolen information on a consumer to successfully impersonate the consumer in credit card transactions, loan applications, and the like.[26]

Both the need for privacy and the need to protect consumers from fraud resulting from identity theft can complicate the cost-benefit trade-off associated with fraud risk. Payments systems that make use of extensive consumer-identifying information can lessen the incidence of fraud, benefiting society. But the value of such information in reducing fraud must be balanced against the value of privacy, and recent cases of identity theft illustrate that such information may not always be used in a socially benevolent fashion.

Conclusion

An important function of any payments medium is to provide certainty of valuation in market exchanges. One of the risks that must be overcome by payments systems is the risk of fraud. Traditional payments media such as currency, checks, and credit cards have effectively contained fraud risk to a level of 20 basis points (0.2 percent) or less. To be successful in the marketplace, newer forms of payment will need to hold fraud risk to similarly low levels.

Incentives for fraud increase when transactions are made in large amounts, when transactions are made anonymously or at the point of sale, when claims cannot be effectively verified at the point of sale, and when issuers of payment claims bear the costs of fraudulent transactions. While these features may be desirable in some situations in that they allow for a greater range of transactions, they can also encourage fraud. The recent Japanese experience with stored-value cards illustrates that vigilance will be necessary in such cases.

Some of the new payments media have been compared with the banknotes used during the U.S. Free Banking Era. The banknotes were subject to substantial fraud risk and were widely discounted. It is unlikely that similar discounting will apply to new payment instruments, however. Modern communications technology and changes in the organization of the banking and payments industries should largely remove incentives for discounting.

Successful payments systems will also have to confront various trade-offs while addressing the problems posed by fraud. These trade-offs include the need to balance the costs of fraud abatement measures with their benefits, the need to balance security of payments systems with consumers' desire for privacy, and the need to encourage development of new, more efficient payments systems while ensuring equitable treatment of participants in these systems.

26. One such identity theft is recounted by Vickers (1996).

REFERENCES

American Bankers Association. 1994. *1994 ABA Check Fraud Survey.* Washington, D.C.

———. 1996. *1996 Bank Card Industry Survey Report.* Washington, D.C.

Bank for International Settlements. 1996a. *Security of Electronic Money.* Basel.

———. 1996b. *Statistics on Payments Systems in the Group of Ten Countries.* Basel.

Bernstein, Nina. 1997. "Online, High-Tech Sleuths Find Private Facts." *New York Times,* September 15, sec. A.

Blinder, Alan S. 1995. Statement before the Subcommittee on Domestic and International Monetary Policy, U.S. House Committee on Banking and Financial Services, October 11.

Bloom, Jennifer Kingson. 1997. "Visa and MasterCard Publish SET Protocol for Internet." *American Banker,* June 5.

Board of Governors of the Federal Reserve System. 1996a. "Electronic Funds Transfers." *Federal Register* 61 (May 2): 19696–705.

———. 1996b. *Report to Congress on Funds Availability Schedules and Check Fraud at Depository Institutions.* Washington, D.C.

———. 1997a. *Report to Congress on the Application of the Electronic Fund Transfer Act to Electronic Stored-Value Products.* Washington, D.C.

———. 1997b. *Report to Congress Concerning the Availability of Consumer Identifying Information and Financial Fraud.* Washington, D.C.

Calomiris, Charles W., and Charles M. Kahn. 1996. "The Efficiency of Self-Regulated Payments Systems: Learning from the Suffolk System." *Journal of Money, Credit, and Banking* 28 (November): 766–97.

Caskey, John P., and Gordon H. Sellon Jr. 1994. "Is the Debit Card Revolution Finally Here?" Federal Reserve Bank of Kansas City *Economic Review* 79 (Fourth Quarter): 79–95.

Congressional Budget Office. 1996. *Emerging Electronic Methods for Making Retail Payments.* Washington, D.C.

Dillistin, William H. 1949. *Banknote Reporters and Counterfeit Detectors, 1826–1866.* New York: American Numismatic Society.

Dwyer, Gerald P., Jr. 1996. "Wildcat Banking, Banking Panics, and Free Banking in the United States." Federal Reserve Bank of Atlanta *Economic Review* 81 (December): 1–20.

Federal Deposit Insurance Corporation. 1996. "Stored-Value Cards and Other Electronic Payments Systems." *Federal Register* 61 (August 2): 40494–97.

Fickenscher, Lisa. 1997. "MasterCard to Cap Consumer Debit Card Liability." *American Banker*, July 31.

Fryer, Bronwyn. 1996. "Visa Cracks Down on Fraud." *Information Week*, August 26.

Glain, Steve, and Norihiko Shirouzu. 1996. "How Japan's Attempt to Slow Nuclear Work in North Korea Failed." *Wall Street Journal*, July 24, sec. A.

Gorton, Gary. 1996. "Reputation Formation in Early Bank Note Markets." *Journal of Political Economy* 104 (April): 346–97.

Greenspan, Alan. 1996. Remarks at the U.S. Treasury Conference on Electronic Money and Banking: The Role of Government, Washington, D.C., September 19.

Hansell, Saul. 1994. "New Breed of Check Forgers Exploits Desktop Publishing." *New York Times*, August 15, sec. A.

Kamihachi, James. 1997. "Supervisory Issues in Electronic Money." Remarks at the *American Banker* Conference on Future Money, June 11. Available on the Internet at http://www.occ.treas.gov/emoney/kami6-11.htm.

Keenan, Charles. 1997. "Visa One-Ups Rival on Consumer Card Liability." *American Banker*, August 14.

Kelley, Edward W., Jr. 1996. Remarks at the CyberPayments '96 Conference, Dallas, Texas, June 18.

Lunt, Penny. 1996. "Is It First and Goal for Debit Cards?" *ABA Banking Journal* 88 (September): 44.

McAndrews, James J. 1997a. "Making Payments on the Internet." Federal Reserve Bank of Philadelphia *Business Review* (January/February): 3–14.

———. 1997b. "Banking and Payments System Stability in an Electronic Money World." Federal Reserve Bank of Philadelphia Working Paper 97-9.

Nielsen, David. 1994. "Check Fraud Rose 136 Percent Over Two Years, ABA Finds." *American Banker*, December 1.

Pearsall, Susan. 1994. "Combating Credit Card Fraud." *New York Times*, December 18, sec. CN.

Pollack, Andrew. 1996. "Counterfeiters of a New Stripe Give Japan One More Worry." *New York Times*, June 20, sec. D.

Rohter, Larry. 1997. "New Bank Fraud Wrinkle in Antigua: Russians on the Internet." *New York Times*, August 20, sec. A.

Rolnick, Arthur J., Bruce Smith, and Warren E. Weber. 1997. "Lessons from a Laissez-Faire Payments System: The Suffolk Banking System (1825–1858)." Federal Reserve Bank of Minneapolis Working Paper 584, September.

Rutledge, Gary. 1996. "Taming the Fraud Monster." *Credit World* (September/October): 10.

Sargent, Thomas J., and Neil Wallace. 1982. "The Real-Bills Doctrine versus the Quantity Theory: A Reconsideration." *Journal of Political Economy* 90 (December): 1212–36.

Schreft, Stacey. 1997. "Looking Forward: The Role for Government in Regulating Electronic Cash." Federal Reserve Bank of Kansas City *Economic Review* 82 (Fourth Quarter): 59–84.

"Survey of Electronic Commerce." 1997. *Economist*, May 10.

Task Force on Stored-Value Cards. 1997. "A Commercial Lawyer's Take on the Electronic Purse: An Analysis of Commercial Law Issues Associated with Stored-Value Cards and Electronic Money." *Business Lawyer* 52 (February): 653–727.

Timberlake, Richard H. 1978. *The Origins of Central Banking in the United States.* Cambridge, Mass.: Harvard University Press.

U.S. Department of the Treasury. 1996. "An Introduction to Electronic Money Issues." Paper prepared for the Treasury Department conference "Toward Electronic Money and Banking: The Role of the Government," Washington, D.C., September 19–20.

U.S. General Accounting Office. 1996. *Counterfeit U.S. Currency Abroad: Issues and U.S. Deterrence Efforts.* Washington, D.C.

———. 1997. *Payments, Clearance, and Settlement: A Guide to the Systems, Risk, and Issues.* Washington, D.C.

Vickers, Marcia. 1996. "Stop Thief! And Give Me Back My Name." *New York Times*, January 28, sec. C.

Weinberg, John A. 1997. "The Organization of Private Payment Networks." Federal Reserve Bank of Richmond *Economic Quarterly* 83 (Spring): 25–43.

White, Eugene N. 1995. "Free Banking, Denominational Restrictions, and Liability Insurance." In *Money and Banking: the American Experience*, 99–118. Fairfax, Va.: George Mason University Press.