Evaluating the Effects of Monetary Policy with Economic Models

TAO ZHA
The author is a senior economist and policy advisor in the macropolicy section of the Atlanta Fed's research department. *He is indebted to Bob Eisenbeis, Frank King, Eric Leeper, Mary Rosenbaum, Ellis Tallman, and Dan Waggoner, whose criticisms and suggestions helped develop this article.*

VOLUMINOUS STUDY OF MACROECONOMICS HAS CONCLUDED THAT MONETARY POLICY OPERATES WITH A LONG LAG. THE LAG WITH WHICH A POLICY CHANGE EXERTS SIGNIFICANT INFLUENCE ON THE ECONOMY IS VARIABLE, DEPENDING ON BOTH PAST POLICY ACTIONS AND THE CURRENT STATE OF THE ECONOMY. IT CAN BE AS LONG AS TWO OR THREE YEARS, ESPECIALLY THE IMPACT ON INFLATION. SUCH A LONG, VARIABLE LAG REQUIRES TODAY'S POLICY ACTIONS TO BE ANTICIPATORY IF THEY ARE TO ACHIEVE A TIMELY IMPACT (KOHN 1995; GREENSPAN 1999; BLINDER 1997; KING 1998). THE POPULAR PRESS AND POLICYMAKERS DESCRIBE THIS FORWARD-LOOKING POLICY AS PREEMPTIVE. IN THE WORDS OF ONE FEDERAL RESERVE GOVERNOR, MONETARY POLICY "IS MOST SUCCESSFUL WHEN IT IS PREEMPTIVE, RESPONDING TO EARLY WARNING SIGNALS OR FORECASTS OF UNFAVORABLE DEVELOPMENTS ON THE INFLATION AND EMPLOYMENT FRONTS" (FERGUSON 1998, 2).

Policymakers always rely on some kind of explicit model to infer the future impact of their actions on economic variables. Such preemptive policy further requires a model to be capable of providing not simply a single-point forecast but, more importantly, a useful evaluation of how alternative policy options might affect the actual economy. By consulting a menu of such alternative outcomes, called policy projections, policymakers then have better information for deciding on a particular policy action that will most likely lead to economic results commensurate with policy objectives.

The task of developing a model to accurately extrapolate quantitative effects of alternative monetary policy actions on the future economy has proved to be a challenge, however. Economists know very well that "a new complete model can easily require years to develop; millions of dollars and careers may be devoted to the effort" (Geweke 1999, 54). It is simply impossible to have a super model that encompasses every aspect of the actual economy. Therefore, a modeler must consciously choose the kinds of questions a model is designed to address. In this article the class of policy questions considered concerns those that policymakers

regularly ask when the central bank is facing a decision about monetary policy. In the U.S. economy, for example, policymakers routinely ask for forecasts of key macroeconomic variables such as inflation and the unemployment rate if the Federal Reserve adopts different federal funds rate paths over the next three years. Assessing a model's usefulness in providing a menu of such projected outcomes for decision making depends on particular criteria. The modeler should therefore be all the more explicit about the criteria set forth in developing the model.

This article focuses on five distinct criteria frequently stressed in the economic literature as important for assessing whether a model is usable. The model should (1) be transparent and reproducible for independent evaluation and further improvement; (2) be able to incorporate new information to update its forecast without ad hoc periodic judgmental adjustments; (3) adequately capture the complicated, dynamic interactions among the multiple key macroeconomic variables of concern to the policymaker; (4) be based on economic theory and offer reasonable economic interpretations of the central bank's behavior; and (5) be able to provide a menu of policy projections under alternative policy scenarios in an economically coherent way. Many existing models meet only some of these criteria. For a model to be usable for actual policy decisions, however, all five are necessary.

The first three emphasize how well an explicit model is fit to the data. The fourth criterion addresses the issue of separating the central bank's behavior from that of the rest of the economy. This issue is referred to as identification of monetary policy, a topic that has received considerable attention in the recent literature (see Zha 1997 and references therein). All four of these criteria are prerequisite to the fifth one—the evaluation of alternative policy scenarios.

This article uses the dynamic, six-variable model of Leeper and Zha (1999) as a pedagogic example of how to combine a well-fit model (criteria 1–3) with a successful identification (criterion 4) to provide a menu of policy projections under alternative policy scenarios (criterion 5). It begins with a discussion of the difficulties and challenges associated with modeling monetary policy in the actual economy. The five criteria are reviewed in light of comparison among different classes of models to highlight the strengths and limitations of empirical modeling for policy projections. The article explains the distinction between a forecasting model and a policy model and then discusses in detail the two distinct aspects of policy making: baseline forecast and identification of monetary policy. The final discussion illustrates the complex assemblage of these two aspects as an integrated process of evaluating monetary policy effects under alternative scenarios.

Difficulties of Modeling Monetary Policy in Reality

To evaluate the effects of monetary policy under alternative scenarios, one must model the behavior of monetary policy. The more explicit and rigorous the model is, the better the model can be understood and improved, and hence result in more effective policy making over time. For the purpose of regular policy making, the model must be able to capture the complex interactions among the key macroeconomic variables that concern policymakers. Because of their complexity, simple rules or expressions, although useful at times, generally do not characterize these interactions adequately.

Because of limited knowledge about how the actual, complex economy operates, policymakers depend on models for understanding the workings of the economy.

In the continuing, day-to-day implementation of U.S. monetary policy, for example, the Federal Reserve constantly evaluates and reevaluates current economic conditions and updates the forecasts of key macroeconomic variables such as inflation and output under alternative policy scenarios. After weighing the alternative outcomes, the Federal Open Market Committee votes on how to direct a main policy instrument, which currently is the federal funds rate. The decision on whether to raise or lower the target for the funds rate or to keep it at the current level depends in large part on the assessment of the dynamic effects of changes in the federal funds rate on various macroeconomic variables such as inflation and the unemployment rate in the next few years.

Such policy making is common across developed countries. At the Bank of France, for example, senior management "assesses the reserve position of the banking system and evaluates whether current market interest rates, especially the interbank rate, are consistent with the current stance of monetary policy and foreign rates. Instructions are then taken to the money market trading room at the Bank of France to intervene in the interbank market on the basis of the evaluations of monetary market and general macroeconomic conditions" (Batten and others 1990, 78). The Bank of Canada "uses economic projections to translate the Bank's objectives into suggested paths for the instruments of policy, and uses various economic and financial indicators, notably monetary aggregates, to monitor progress and help the Bank to act in a timely fashion when necessary" (Duguay and Poloz 1994, 197).

To what extent a particular monetary policy action will achieve its intended effects on output growth and inflation depends crucially on how the economy develops over time. Many unforeseen developments are outside the central bank's control: an unexpected decline in commodity prices, a surprising improvement in the labor market, an unanticipated strength in output growth, or a sudden fall in the exchange rate, for example. These private shocks, coupled with the central bank's particular policy actions, affect economic performance and often drive the economy along paths different from those forecast. When favorable or unfavorable shocks arrive, policymakers need to "be flexible in revising forecasts and the policy stance in response to new information contradicting their previous predictions" (Kohn 1995, 235) and adopt different policy actions accordingly (Blinder 1997).

Consumers and producers in the private sector also evaluate the dynamic impact of monetary policy regularly in making their own investment decisions. They understand that particular policy actions depend on the changing state of the economy. Thus their investment plans take into account the uncertain factors in actual monetary policy actions. The outcome of such a complex interplay between monetary policy and the private sector is reflected in the data observed in the actual economy. The data themselves, however, do not distinguish the behavior of the private sector from that of the central bank and depend on models to infer what the policy behavior is. Because different models lead to different conclusions, learning the central bank's behavior from the data requires careful effort in approximating the actual economy in a workable framework.

The notion of "workable" is important because there does not and cannot exist a model that perfectly represents the actual economy. The central bank's real-life behavior, as described above, is far more complicated than can be completely captured by any kind of model, empirical or theoretical. All well-specified models at best approximate the actual economy, offering different perspectives on the key interactions among a set of variables and having their relative strengths in some dimensions and weaknesses in others. Thus, the reasonableness of a particular model depends on how usable the model is for certain questions and under certain criteria.

Three Classes of Models

The policy questions addressed in this article concern evaluating the dynamic effects of monetary policy as the central bank faces decisions about raising or lowering the interest rate. This section analyzes three popular classes of models used for this kind of policy analysis.

One class of models searches for a *single policy variable* (such as a monetary aggregate or an interest rate) as an indicator of monetary policy. According to this approach, the policy variable, be it a money stock or an interest rate, is controlled by the central bank but unaffected by other variables. Policy actions must evolve autonomously, independent of the changing state of the economy. This scenario allows tracing the effects of alternative policy choices represented by the movements in the policy variable conveniently and unambiguously through the variable's correlations with other macroeconomic variables.

While such a model offers simplicity, its conditions are scarcely met in actual economies. It would therefore be difficult to have economically coherent interpretations for imposing these conditions in a policy model (Tobin 1970). For example, the federal funds rate target is not set autonomously; rather, it is regularly adjusted to reflect the Federal Reserve's concern about its own objectives of, for instance, price stability and full employment. When fluctuations in economic activity or the repercussions of past policy choices threaten this objective under the current level of the federal funds rate, a new target rate will be chosen. Clearly, there is feedback between the state of the economy and the policy variable. The practical reality of state-dependent policy choices makes it most likely that the conditions underlying the indicator approach violate criteria 4 and 5.

Another class of models seeks a *simple rule* for describing the central bank's behavior. Simple rules provide a convenient or even compelling way for policy analysts to explain complex economic activity to policymakers, but they are unlikely to adequately capture the intricate dynamics taking place in the actual economy. Consequently, the assumptions embedded in these rules are often questionable. The assumption of a NAIRU (nonaccelerating inflation rate of unemployment) relationship is an example. This rule states that whenever the unemployment rate is below some threshold level, inflation will soon rise. If such a threshold level could be unambiguously determined and if such a relationship were stable, the rule would provide an appealing story to policymakers about the future path of inflation. The threshold level of unemployment, however, cannot, in fact, be measured. It is often estimated with large errors (Staiger, Stock, and Watson 1997). The estimation is fragile because this rule ignores the effects of other important factors (such as monetary policy itself) on the path of inflation and employment (Chang 1997; Espinosa and Russell 1997). Consequently, this rule provides neither a forecast of macroeconomic variables other than inflation nor a forecast of inflation under a different policy option, so it does not meet criteria 3 and 5.

Another example of simple rules is the Taylor rule. In its often-used version the Taylor rule states that the federal funds rate changes in response to only two variables: the gap between current actual gross domestic product (GDP) and potential GDP and the four-quarter inflation rate. The attractiveness of this rule is its ability to present a simple story involving only these two variables. Such simplicity, however, has serious weaknesses of its own. Because potential GDP is an abstract concept rather than something that can be measured, the movement in the fed funds rate crucially depends on how potential GDP is estimated. Like determining the threshold of the unemployment rate in the NAIRU rule, the estimation of potential GDP can be very imprecise and controversial. Furthermore, in estimating the Taylor rule it is often assumed that movements in output and inflation are independent of those in the federal funds rate. This assumption itself disables the rule from assessing the dynamic effects on output and inflation of changes in the federal funds rate (criteria 3 and 5).

In some research programs, the Taylor rule is used as one of many relationships in a larger model. In such a model, changes in the Taylor rule are examined and the effects on macroeconomic variables of these changes are analyzed. In essence, the changes arbitrarily alter the way and the degree in which the fed funds rate responds to the GDP gap and inflation. These exercises are often undertaken under the assumption that a change in the Taylor rule does not affect the observed relationships among a set of macroeconomic variables in the actual economy. But this assumption is economically incoherent because, as Lucas (1976) has long argued, the observed relationships among macroeconomic variables will change with different policy rules. Because of the Lucas critique, as this argument is dubbed in the economic literature, these kinds of exercises examining monetary policy effects do not meet criterion 4.

A third class of models comprises *large-scale econometric models* designed to capture in detail the structure of the real economy as completely and accurately as possible. Such a model often deals with a large number of industries and sectors of the economy and may involve hundreds of equations and variables. The main objective of most large-scale models is not simply to provide forecasts of multiple key macroeconomic variables. Rather, it is to provide detailed stories about the economy and to assess the impacts of different kinds of shocks at a disaggregate level unavailable in smaller models. Policymakers would like to be informed about the details of the state of the real economy. They may want to know, for instance, what happens to the durable goods sector, what the outlook is for the labor market in the service sector, or what the impact is on the U.S. economy of the Asian financial crisis.

Large-scale models, however, also come with costs. They are often difficult to reproduce and evaluate independently (criterion 1). Because modifying so large a model can be quite costly, judgmental adjustments are called in periodically to address new, unanticipated information (criterion 2). Furthermore, given a short span of historical data observed in the actual economy, it is impossible to have precise estimation of hundreds and thousands of equations and variables in a single framework. Thus, the large-scale model is often broken into separate parts by imposing strong assumptions. Many of these assumptions have been criticized by Sims (1980) as "incredible" because they are imposed not from the viewpoint of having a reasonable approximation of the economy as a whole but from the separate, partial consideration of keeping various parts of the model manageable (criterion 3). These drawbacks are likely to compromise the modeler's original aim of representing the detailed structure of the economy as completely and accurately as possible.

For models to be usable for evaluating monetary policy effects, modelers must recognize that fluctuations or shocks in the actual economy are often driven by developments beyond the central bank's control.

Dynamic Multivariate Modeling

In this and subsequent sections, the discussion turns to another class of models—dynamic multivariate models—and explains their advantages in light of the five criteria set forth in the introduction.[1] The term *dynamic* connotes the idea that economic variables influence one another through variable lags over time. For example, a change in the interest rate today

1. *Dynamic multivariate models are frequently referred to as vector autoregressive (VAR) models in the economic literature. A sample of the work includes Bernanke (1986), Blanchard and Watson (1986), Sims (1986), Bernanke and Blinder (1992), Gali (1992), Gordon and Leeper (1994), Strongin (1995), Pagan and Robertson (1995), Cochrane (1996), Leeper, Sims, and Zha (1996), Cushman and Zha (1997), Bernanke, Gertler, and Watson (1997), Christiano, Eichenbaum, and Evans (forthcoming), Bernanke and Mihov (1998), and Robertson and Tallman (1999a, 1999b).*

will affect inflation over the next few years. The term *multivariate* implies a single framework in which multiple economic variables are considered. Thus, a dynamic multivariate model is a single framework that uses multiple equations to incorporate the dynamic relationships among multiple economic variables.

Dynamic multivariate modeling offers an approach to policy analysis that is different from those discussed in the previous section. It is designed to address a small set of recurring questions that constitutes the core of policymakers' concerns. Policymakers need to know, on a regular basis, how the future paths of key macroeconomic variables such as output, inflation, and the unemployment rate will change if the policy instrument—in this case the federal funds rate—follows different paths in the future. The main advantage of dynamic multivariate modeling is to focus on this set of questions by evaluating, as accurately as possible, the quantitative effects of policy actions on key macroeconomic variables under different policy scenarios. To this end, a system of multiple equations in a dynamic multivariate model avoids postulating a simple but unrealistic rule of monetary policy. Such a system approach explicitly recognizes the intertwined, complex relationships between a policy variable like the interest rate and other key macroeconomic variables such as inflation and the unemployment rate (Leeper, Sims, and Zha 1996). At the same time, the dynamic multivariate model typically selects only a small set of key macroeconomic variables to avoid unreasonable assumptions in the model's estimation. In other words, dynamic multivariate modeling is designed to capture the joint, complicated behavior of the central bank and the private sector in one single framework without imposing too many strong assumptions and without sacrificing its connection to modern economic theories (Leeper and Sims 1994; Sims 1996; Diebold 1998; Cooley and Quadrini 1998a, 1998b). In the discussion below, the Leeper-Zha model for the U.S. economy illustrates how a dynamic multivariate model is designed to meet the first four of the five criteria listed in the introduction section.

Baseline Forecast. When a dynamic multivariate model is constructed, a minimum set of restrictions needs to be imposed. These restrictions include the choice of variables, the length of time over which variables are allowed to interact with one another, and some mathematical assumptions that make the model tractable (criterion 1). As a starting point, the model avoids further restrictions such as particular economic views on the exact interactions between monetary policy and the private sector. Because these economic views are not present, such a model is called in technical parlance a reduced-form model. The reduced-form model is designed to allow the data, not the modeler, to determine the complicated dynamic interactions among the economic variables. Thus, the model is designed to fit to the data (criterion 3).

A baseline forecast, sometimes referred to as a reduced-form forecast, is produced from a reduced-form model. Once the model is specified, the baseline forecast is updated upon the arrival of new information without ad hoc adjustments (criterion 2). Often, the forecasting performance is measured by the difference between the baseline forecast and the actual outcome over time. The performance thus measured has sometimes been compared with other forecasts, and evidence has shown that it has been comparable. (Litterman 1986; McNees 1986; Meyer 1998; Zha 1998; Robertson and Tallman 1999a).

The reduced-form Leeper-Zha model is a dynamic, six-variable monthly model consisting of a system of six equations. Besides an index of commodity prices, the variables include the key macroeconomic variables policymakers are most concerned about: the federal funds rate, the M2 stock, the consumer price index (CPI), real (inflation-adjusted) GDP, and the unemployment rate. The six equations, as a system, allow policy variables such as the federal funds rate to interact with other macroeconomic variables such as output and the CPI, both within a month and through variable lags (criterion 3). Because of the complicated dynamics inherent among actual macroeconomic variables, this dynamic and multiple-equation feature of the model is critical for allowing the variables to interact with one another contemporaneously and over time without ad hoc periodic adjustments (criterion 2).

The movements in these macroeconomic variables tend to be very persistent over time. Thus, when modeled in one framework, the past values of the variables often have strong predictive power in forecasting future values. As a result of this feature of persistence, the Leeper-Zha model has consistently produced baseline forecasts of these key macroeconomic variables over the past twenty years that are comparable to other forecasts. The fact that the model is small-scale and explicit makes it transparent enough to be reproduced, evaluated, and improved over time (criterion 1).

Although it is tempting to add more variables to such a small-scale model, the addition would have costs. Either it is increasingly infeasible to obtain precise estimation with the model as it grows larger or one must impose ad hoc strong assumptions or make judgmental adjustments to keep the size of the model manageable. These costs are not trivial as they often lead to conflicting economic stories and even misleading policy analysis.

While it is true that the Federal Reserve has information about hundreds and even thousands of variables beyond the six key macroeconomic variables in the Leeper-Zha model, the issue is not whether the Federal Reserve has the data for a large number of variables.

Instead, the real issue is related to criterion 3. That is, the issue is whether other variables (for example, the number of orders for automobiles, workers' wage compensations, or the consumer confidence index) would significantly help in modeling the interactions among the variables the Federal Reserve is ultimately interested in. No consistent evidence indicates that adding other variables would help the model better fit those core macroeconomic variables.

Identification of Monetary Policy. The foregoing subsection discusses the reduced-form Leeper-Zha model in light of criteria 1–3. The reduced-form model is often used purely as a forecasting tool in the forecasting literature, so the performance of baseline forecasts in comparison with others has been a central focus. There is a tendency, however, to overemphasize forecasting performance as a sole criterion in judging whether the model can be used in evaluating monetary policy actions. Doing so can be seriously misleading.

To understand this argument it is important to note the distinction between a forecasting (reduced-form) model and a policy (structural) model, which has profound implications about the importance of an identification of monetary policy (criterion 4). The reason that a forecasting model is regarded as a reduced form rather than a structural form is that, as argued before, it imposes no economic structure to distinguish the central bank's behavior from the rest of the economy. Thus, the scenario represented by a baseline forecast is often not the one that interests policymakers. The baseline forecast serves as a basis only in the technical sense that the modeler is able to conveniently use it to produce a menu of alternative scenarios in which policymakers may be interested. At a minimum these alternative scenarios require imposing restrictions on the interactions between economic variables that allow extracting the central bank's behavior from the data. These restrictions are called identifying restrictions, and this process of sorting out the central bank's behavior from that of the rest of the economy is, as explained earlier, called identification of monetary policy. A model that is able to identify monetary policy is called a structural model or policy model. Clearly, the structural or policy model imposes more restrictions than the reduced-form model.

Because the observed data are the outcome of the dynamic, complex interplay between monetary policy and the private sector, they themselves do not distinguish the behavior of the Federal Reserve from that of the rest of the economy. Therefore, in addition to the minimum set of restrictions imbedded in the reduced-form model, one must further impose a particular economic view on the model in order to identify monetary policy. Any set of identifying restrictions used to reflect such a view can be controversial because economists disagree on particular views about how the actual economy works. Such dissension largely stems from the complicated nature of the economy, of which economists have limited understanding and which no single model can encompass. The fact that identifying restrictions can be controversial, however, by no means implies that one should abandon developing a formal (that is, model-based) economic framework. Rather, it means that when particular identifying restrictions are imposed, the economic meanings behind these restrictions must be explained carefully and explicitly in the context of a model.

It is simply impossible to have a super model that encompasses every aspect of the economy. Therefore, a modeler must consciously choose the kinds of questions a model is designed to address.

The phrase "in the context" is the quintessence of credible identification because without an explicit model to serve as a framework it would be impossible to distinguish the Federal Reserve's own behavior from that of the rest of the economy. And if the Federal Reserve's behavior is not explicitly specified, there is no way to evaluate the quantitative effects of different policy actions in a formal, transparent way (criterion 1). In other words, an explicit model provides a context in which the effects of monetary policy can be quantified and evaluated (Shapiro 1994).

Although a particular set of identifying restrictions may not be accepted universally, it should be guided by economic theory and have reasonable economic interpretations (criterion 4). A previous article in this publication by Zha (1997) discusses this issue in detail. Here, the meaning of criterion 4 is illustrated in the context of the Leeper-Zha model.

Identification in the Leeper-Zha model is accomplished by specifying interactions among macroeconomic variables in terms of several sectors. One of the sectors is the money market. In that market, both the demand for and supply of money determine the level of the interest rate. When the demand for money increases, the interest rate is driven up. If the Federal Reserve desires to keep the interest rate from rising, it must supply more reserves, thereby causing an increase in broad money stock (here, M2). Thus, one equation in the model, called the money demand equation (MD), describes the behavior of money demand, and one equation, called the monetary policy

equation (MP), describes how the Federal Reserve supplies money to keep the interest rate at a certain level:

$$M = \alpha_1 R + \alpha_2 P + \alpha_3 y + \alpha_4 X_{MD} + \varepsilon_{MD} \text{ (MD)}$$
$$R = \beta_1 M + \beta_2 P_{cm} + \beta_3 X_{MP} + \varepsilon_{MP} \quad \text{(MP)}. \tag{1}$$

In system (1), M stands for M2; R, for the federal funds rate; P, the CPI; P_{cm}, the index of commodity prices; y, real GDP; X_{MD}, a set of lagged variables in the MD equation; and X_{MP}, a set of lagged variables in the MP equation.[2] Lagged variables are important in helping predict the dynamic fluctuations of the macroeconomic variables in the future. The values of the α parameters in the MD equation and the β parameters in the MP equation, unlike those of the variables, do not come from the data and thus must be estimated. The notation ε_{MD} denotes the money demand shock, and ε_{MP}, the monetary policy shock. The term *shock* is used because it describes behavior that cannot be predicted by the model, a point that will be further discussed in the next section.

The two equations have reasonable economic interpretations by standard economic theory: In the MD equation in system (1), the demand for money depends on income (which is approximated by real GDP), the interest rate, and the price level. The MP equation implies that the Federal Reserve can change the interest rate by influencing the money stock and by quickly responding to changes in the index of commodity prices (which serve as a signal of future inflation). Because of the delay in the release of the data on output and the price level, however, the Federal Reserve cannot respond to changes in these variables instantly (here, within a month). But the Federal Reserve can respond to changes in the lagged variables (X_{MP}), which serve to predict the current and future fluctuations of output and CPI. Because M and R enter both equations in system (1), the money stock (M) and the interest rate (R) are determined by both the MD and MP equations simultaneously.

The successful identification of monetary policy involves estimating all parameters jointly from the data. This joint estimation and inferential conclusions drawn from this estimation present a technical challenge.[3] Nonetheless, the joint feature of the model as shown in system (1) is needed to realistically account for the simultaneous and dynamic interactions among policy variables and other macroeconomic variables. This realistic account is an important aspect of reasonableness in economic interpretations (criterion 4). In the money market, for instance, it is known that the money stock and the interest rate influence each other simultaneously through both demand and supply of money. The results would be misleading if one assumes away this simultaneous interaction without having confirmation by empirical estimates of the parameters α_1 and β_1 in system (1). If, for example, β_1 is assumed to be zero but the empirical estimate of β_1 turns out to be significantly different from zero, the monetary policy equation would be misspecified by the zero β_1 assumption.

A Menu of Policy Projections under Alternative Scenarios

The preceding sections discuss the reduced-form Leeper-Zha model according to criteria 1–3 and the structural Leeper-Zha model according to criterion 4. This section offers an intuitive explanation of how such a structural model can be used to produce a menu of policy projections under alternative policy options. The discussion addresses criterion 5, that is, the model's ability to provide alternative forecasts under different policy scenarios. It is this ability that marks structural dynamic multivariate models as promising and rich tools in evaluating the effects of monetary policy in an economically coherent way (without violating the Lucas critique).

To demonstrate how the structural Leeper-Zha model—the policy model—can be used to provide alternative forecasts under different policy options, the discussion begins with technical notions of the terms *endogenous* and *exogenous*. The part that can be predicted by the model—for example, the terms with the α and β parameters in system (1)—is endogenous to the model; the part that cannot be predicted by the model is exogenous to the model—for example, ε_{MP} and ε_{MD}—and thus is often approximated by a stochastic (random) process. Because of this random feature, ε_{MP} is called a monetary policy shock and ε_{MD} a money demand shock. Clearly, the shock ε_{MP} makes sense only in the context of a specific model. What can be predicted by one model may not be consistent with what another predicts. What is a random shock to a particular model may not be random or exogenous to other models or from the perspective of particular policymakers.[4]

In the context of the Leeper-Zha model, a menu of policy options and their corresponding effects can be produced by combining the baseline forecast and exogenous shocks. The baseline forecast is a projection under the assumption that there is no shock in the economy (that is, $\varepsilon_{MP} = \varepsilon_{MD} = 0$). This scenario seldom occurs because there will always be shocks in the future. A projection that deviates from the baseline reflects a scenario in which the effect of monetary policy is different from what the baseline implies. Such a projection therefore combines the baseline with a hypothetical path of exogenous shocks.

Charts 1 and 2 depict the MP and MD equations in system (1) on the R and M plane. These two charts provide examples of two simple scenarios in which alternative policy projections are simulated with the policy model. In both charts the horizontal axis represents the money stock, M, and the vertical one represents the interest rate, R. In Chart 1, the two thick red lines represent the baseline situation in which $\hat{P}_{cm}$, $\hat{P}$, $\hat{y}$, $\hat{R}$, and $\hat{M}$ represent the baseline forecast and X_{MP} and X_{MD} are the data that have been observed. The thick red MP line, as indicated in system (1), depends on $\hat{P}_{cm}$, X_{MP}, and ε_{MP} (whose value here is zero). Similarly, the thick red MD line depends on $\hat{P}$, $\hat{y}$, X_{MD}, and ε_{MD} (whose value is also zero). These two red lines intersect at $\hat{R}$ and $\hat{M}$ as an equilibrium outcome. The forecast variables $\hat{P}_{cm}$, $\hat{P}$, $\hat{y}$, and $\hat{M}$ represent the effect of monetary policy consistent with the funds rate, $\hat{R}$. If, for example, policymakers want to explore the effect on the macroeconomic variables of lowering the funds rate to the level of R^*, one can use the model to compute how much an exogenous shift in monetary policy is required to achieve this target. If the computed value of such a shift is ε^*_{MP}, one can calculate a new forecast, denoted as, P^*_{cm}, y^*, R^*, and M^*. Consequently, the two red lines move to the positions of the two thin black lines. Thus, the new forecasts, P^*_{cm} (commodity prices index), P^* (the general price level), y^* (GDP), and M^* (M2), different from the baseline, represent the effect of this new policy choice of keeping the federal funds rate at R^*.

Chart 2 presents a more complicated situation. Suppose that the Federal Reserve has information about a liquidity problem in the banking system. Furthermore, suppose that this information is not captured by the predictable (endogenous) part of the model but can be approximated by the random (exogenous) part in the money demand equation. If the value of this exogenous MD shock is $\tilde{\varepsilon}_{MD}$, the model's forecast will deviate from the baseline. This deviated forecast is denoted by $\tilde{P}_{cm}$, $\tilde{P}$, $\tilde{y}$, $\tilde{R}$, and $\tilde{M}$. The two thick red lines in Chart 2 represent the MP and MD equations in this situation. Policymakers may also be concerned about the inflation rate implied by the price level, $\tilde{P}$, and want to bring the inflation rate down. If $\bar{P}$ represents the price level consistent with the inflation level desired by the policymakers, the model can then be used to calculate the corresponding monetary policy shock, the value of which is denoted by $\bar{\varepsilon}_{MP}$. The new forecast of other macroeconomic variables can also be simulated through the model, with the value of this new forecast denoted by $\bar{P}_{cm}$, $\bar{P}$, $\bar{y}$, $\bar{R}$, and $\bar{M}$. Graphically, the two red lines in Chart 2 move to the two thin black lines, the intersection of which gives $\bar{R}$ and $\bar{M}$. Chart 2 shows that the model can be used to advise policymakers about not only the level of the interest rate, ($\bar{R}$), they need to target to achieve the price level, $\bar{P}$, but also the forecast of other macroeconomic variables.

CHART 1
Policy Change to Target the Funds Rate

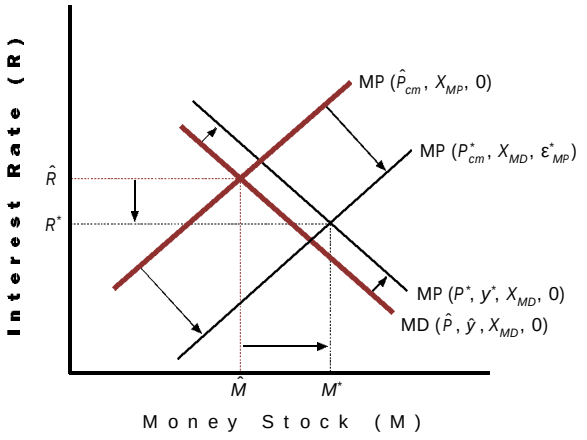
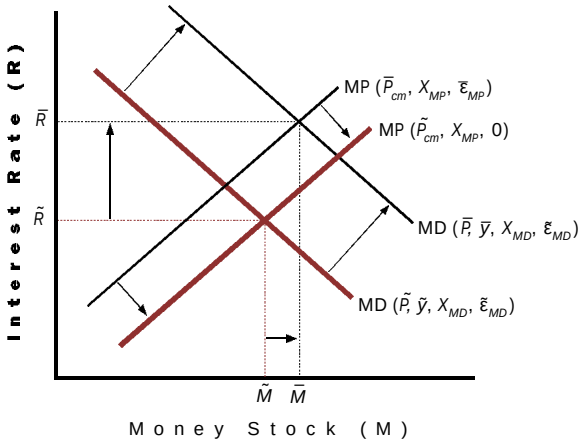


CHART 2
Policy Change to Target the Price Level



2. *All variables except for the interest rate and the unemployment rate are logarithmic.*
3. *See Waggoner and Zha (1999) for details.*
4. *This simple point is not only common across all disciplines in economics but also valid in other social sciences. As novelist Tony Hillerman has amusingly but poignantly observed, "From where we stand, the rain seems random. If we could stand somewhere else, we would see the order in it" (quoted in Robert 1994).*

The charts are used for illustration only. The use of the identified dynamic multivariate model for actual policy projections is far more complicated because it involves not only the dynamic paths of macroeconomic variables under a particular policy option but also the uncertainty around these paths into the future. Nonetheless, Charts 1 and 2 illustrate a point that is central to the approach proposed in Leeper and Zha (1999). The point is that exogenous shocks, such as the MD and MP shocks, are relevant only in connection to the baseline forecast. In other words, the dynamic patterns of exogenous shocks and the baseline forecast constitute an integrated process for quantifying different forecasts of key macroeconomic variables if the federal funds rate follows alternative paths. Such a menu of various policy projections is at the heart of the use of dynamic multivariate modeling for advising policymakers.

Model Improvement

The discussion in this section focuses on what has been learned from the dynamic multivariate model and in what directions the current dynamic multivariate modeling can be improved. The dynamic multivariate model offers a tool for decomposing the model into endogenous and exogenous elements. This decomposition is vitally important for simulating policy projections in an economically coherent way (criterion 5), meaning that the values of parameters (such as α's and β's) do not vary with exogenous shifts in monetary policy. This invariance is the essence of the Lucas critique, which cautions the use of an empirical model if the values of parameters in such a model change with policy shifts. Although discussion of the Lucas critique is beyond the scope of this article, it is important to note that the critique is a subtle concept, and the previous use of dynamic multivariate models has been problematic in this regard.[5]

It is also important to point out that the separation of what is endogenous and what is exogenous is a convenient byproduct of the model, which offers a particular way to help modelers examine and understand the economy. Any model, whether it is the dynamic multivariate model discussed above or another kind, is only an imperfect abstraction of the complex real economy. Different economists or modelers may have different specifications of the Federal Reserve's behavior and the behavior of the private sector, and what is exogenous to one modeler may not be to others. The comparison of exogenous components across different models, as some recent literature has attempted (for example, Rudebusch 1998), misses this fundamental point. The key insight gained so far leads to a need to combine, not separate, endogenous and exogenous movements. It is this combination that provides a way for modelers to produce projections of policy effects under alternative scenarios. Perhaps there are other economically coherent ways of making similar projections. Ultimately it is the accuracy of such projections that matters to policymakers in their policy decisions. While the model must make both economic and mathematical sense, the goal of model improvement should aim at raising the accuracy of policy projections.

Broadly speaking, the current dynamic multivariate approach can be improved in two directions. First, the dynamic multivariate model could be more closely connected to modern economic theories as called for by Ingram and Whiteman (1994), Leeper and Sims (1994), Sims (1996), and Diebold (1998). The dynamic multivariate model would then be able to offer more detailed economic interpretations or stories for policymakers. Currently, however, the conceptual and technical difficulties associated with such a connection in a multiple-equation framework are overwhelming.[6] Finding a particular way of connecting theory to the dynamic multivariate model that improves, not impairs, the forecasting accuracy will be a long and incremental process.

In the other direction, the empirical features in a class of current dynamic multivariate models could be further refined. The current dynamic multivariate approach often maintains the assumption that the structure of the economy is linear and that the Federal Reserve's behavior remains more or less stable across time. In technical terms these features are called linearity and time-invariance. Researchers have realized the need to relax these features by allowing some kind of nonlinearity and time-variation in the dynamic multivariate model (Sims 1993), but the challenge is to determine what kind of nonlinearity and time-variation would improve current dynamic multivariate models. Most current time-variation and nonlinearity literature has focused exclusively on univariate cases rather than the multivariate framework that is crucial for policy analysis. Furthermore, these works maintain the strong assumption that the real economy evolves in sudden, exogenous, and disruptive ways. Empirical evidence has not supported such an assumption (Sims 1993; Zha 1998). In fact, the evidence in the economic literature has shown little advantage, if not much disadvantage, of nonlinear and time-varying multivariate models (Sims 1993; Uhlig 1997).

The issue here is not whether the economy evolves in a nonlinear and time-varying way because clearly it does. The real issue is how to depict these characteristics in ways that would best characterize the economy and whether researchers have the tools to handle nonlinear, time-varying multivariate models. If introduced inappropriately, nonlinearity and time-variation in a model could deliver a worse forecast than linear multivariate models.

A recent study by Harding and Pagan (1998) offers an illuminating example. Harding and Pagan use different kinds of nonlinear and time-varying models in the existing literature to characterize the business cycles of the real economy. What they find is that the simple linear model (the unit-root model in technical terms) dominates all other seemingly sophisticated models in describing the pattern of business cycles in both the United States and other developed countries. The point, already made, is not that nonlinear and time-varying models are inferior but that introducing a kind of nonlinearity and time variation that can better approximate the real economy is not nearly so straightforward as it may seem.

Clearly, developing model improvements to ensure more accurate results is challenging. Since the economy changes gradually in degrees that are unknown to researchers, determining how to connect the dynamic multivariate model to economic theory or to introduce some kind of nonlinearity and time variation in the dynamic multivariate model requires all the more deliberate and careful effort. Meanwhile, researchers and policy analysts have learned that the linear and time-invariant dynamic multivariate framework provides a reasonable approximation to the economy in comparison with other existing types of models. The assumption of linearity and time invariance allows researchers to overcome some technical hurdles otherwise associated with dynamic multivariate models and to gain a deeper understanding of both the strengths and limitations of these kinds of models (Sims and Zha 1998; Waggoner and Zha 1998, 1999). Such understanding is a necessary step in exploring the feasibility and capacity for improving the dynamic multivariate model or perhaps even replacing it with a viable alternative.

Conclusion

Because of limited knowledge about how the actual, complex economy operates, policymakers depend on models for understanding the workings of the economy. For models to be usable for evaluating monetary policy effects, modelers must recognize that fluctuations or shocks in the actual economy are often driven by developments beyond the central bank's control. There are no simple rules, and neither is there a single model that represents the exact interactions between monetary policy and the rest of the economy (Duguay and Poloz 1994). How good a model is depends on particular criteria. Therefore, modelers must specify a set of criteria in constructing a particular model.

This article assesses the usability of a dynamic multivariate model for policy evaluation on the basis of five specific criteria. In summary, (1) the model's small scale enables the modeler to understand the specific dimensions along which the model can be improved; (2) the single framework enables the model to update the forecast without ad hoc periodic judgmental adjustments; (3) the dynamic, multiple-equation nature of the model enables it to provide a reliable forecast of multiple key macroeconomic variables; (4) the identifying restrictions imposed on the model enable the modeler to infer the central bank's behavior from the data; and (5) a menu of policy projections under alternative scenarios produced from the model provides useful guidance for preemptive monetary policy.

There are no simple rules, and neither is there a single model that represents the exact interactions between monetary policy and the rest of the economy.

With these five criteria in mind, this article uses the example of the Leeper-Zha model to address a set of recurring practical questions regularly asked by policymakers. These questions concern projecting multiple key macroeconomic variables under alternative policy scenarios at the time when the policy decision has to be made. The discussion focuses on the two conceptual issues that are central to answering these questions: the baseline forecast in reduced-form models and identification in structural-form models. The article explains the distinction between forecasting (reduced-form) and policy (structural) models. The use of a baseline forecast serves only as a convenient technical tool for computing a menu of policy projections under alternative scenarios. The important message is that an assemblage, not a separation, of baseline forecast and identified policy shifts provides economically coherent ways of evaluating the effects of monetary policy.

5. See Leeper and Zha (1999) for detailed discussion.
6. See, for example, Leeper and Sims (1994) and Waggoner and Zha (1998, 1999).

REFERENCES

BATTEN, DALLAS S., MICHAEL P. BLACKWELL, IN-SU KIM, SIMON E. NOCERA, AND YUZUVU OZEKI. 1990. "The Conduct of Monetary Policy in the Major Industrial Countries: Instruments and Operating Procedures." International Monetary Fund Occasional Paper 70, July.

BERNANKE, BEN S. 1986. "Alternative Exploration of the Money-Income Correlation." *Carnegie-Rochester Conference Series on Public Policy* 25 (Autumn): 49–99.

BERNANKE, BEN S., AND ALAN S. BLINDER. 1992. "The Federal Funds Rate and the Channels of Monetary Transmission." *American Economic Review* 82 (September): 901–21.

BERNANKE, BEN S., MARK GERTLER, AND MARK WATSON. 1997. "Systematic Monetary Policy and the Effects of Oil Price Shocks." *Brookings Papers on Economic Activity*, no. 1: 91–142.

BERNANKE, BEN, AND ILIAN MIHOV. 1998. "Measuring Monetary Policy." *The Quarterly Journal of Economics* (August): 869–902.

BLANCHARD, OLIVIER JEAN, AND MARK WATSON. 1986. "Are Business Cycles All Alike?" In *The American Business Cycle: Continuity and Change*, edited by Robert Gordon. Chicago: University of Chicago Press.

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

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

CHRISTIANO, LAWRENCE J., MARTY EICHENBAUM, AND CHARLES EVANS. Forthcoming. "Monetary Policy Shocks: What Have We Learned and to What End?" In *The Handbook of Macroeconomics*, edited by John Taylor and Michael Woodford.

COCHRANE, JOHN H. 1996. "What Do the VARs Mean? Measuring the Output Effects of Monetary Policy." University of Chicago. Unpublished paper.

COOLEY, THOMAS F., AND VINCENZO QUADRINI. 1998a. "Monetary Policy and the Financial Decisions of Firms." University of Rochester and Duke University. Unpublished paper.

———. 1998b. "A Neoclassical Model of the Phillips Curve Relation." University of Rochester and Duke University. Unpublished paper.

CUSHMAN, DAVID O., AND TAO ZHA. 1997. "Identifying Monetary Policy in a Small Open Economy under Flexible Exchange Rates." *Journal of Monetary Economics* 39 (August): 433–48.

DIEBOLD, FRANCIS X. 1998. "The Past, Present, and Future of Macroeconomic Forecasting." *Journal of Economic Perspectives* 12 (Spring): 175–92.

DUGUAY, PIERRE, AND STEPHEN POLOZ. 1994. "The Role of Economic Projections in Canadian Monetary Policy Formulation." *Canadian Public Policy* 20 (June): 189–99.

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.

FERGUSON, ROGER W., JR. 1998. *Exercising Caution and Vigilance in Monetary Policy*. Speech to the Federal Reserve Bank of Atlanta, July 9.

GALI, JORDI. 1992. "How Well Does the IS-LM Model Fit Postwar U.S. Data?" *Quarterly Journal of Economics* 107 (May): 709–38.

GEWEKE, JOHN. 1999. "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communications." *Econometric Reviews* 18 (February): 1–73.

GORDON, DAVID B., AND ERIC M. LEEPER. 1994. "The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification." *Journal of Political Economy* 102 (December): 1228–47.

GREENSPAN, ALAN. 1999. Statement before the House Subcommittee on Domestic and International Monetary Policy, Committee on Banking and Financial Services. 106th Cong., 1st sess. July 22.

HARDING, DON, AND ADRIAN PAGAN. 1998. "Knowing the Cycle." University of Melbourne and Australian National University. Unpublished paper.

INGRAM, BETH F., AND CHARLES H. WHITEMAN. 1994. "Supplanting the 'Minnesota' Prior Forecasting Macroeconomic Time Series Using Real Business Cycle Model Priors." *Journal of Monetary Economics* 34 (December): 497–510.

KING, MERVYN. 1998. "The UK Economy and Monetary Policy—Looking Ahead." *Bank of England Quarterly Bulletin* (August): 284–86.

KOHN, DONALD L. 1995. "Comment on 'Inflation Indicators and Inflation Policy' by Cecchetti." *NBER Macroeconomics Annual, 1995*, edited by Ben S. Bernanke and Julio J. Rotemberg. Cambridge: MIT Press.

LEEPER, ERIC M., AND CHRISTOPHER A. SIMS. 1994. "Toward a Modern Macroeconomic Model Usable for Policy Analysis." In *NBER Macroeconomics Annual, 1994*, edited by Stanley Fischer and Julio J. Rotemberg. Cambridge: MIT Press.

LEEPER, ERIC M., CHRISTOPHER A. SIMS, AND TAO ZHA. 1996. "What Does Monetary Policy Do?" *Brookings Papers on Economic Activity*, no. 2:1–63.

LEEPER, ERIC M., AND TAO ZHA. 1999. "Modest Policy Interventions." Federal Reserve Bank of Atlanta Working Paper 99-22, December.

LITTERMAN, ROBERT B. 1986. "Forecasting with Bayesian Vector Autoregressions—Five Years of Experience." *Journal of Business and Economic Statistics* 4 (January): 25–38.

LUCAS, ROBERT E., JR. 1976. "Econometric Policy Evaluations: A Critique." In *The Phillips Curve and Labor Markets*, edited by Karl Brunner and Allan H. Meltzer. Amsterdam: North-Holland.

MCNEES, STEPHEN K.1986. "Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts." *Journal of Business and Economic Statistics* 4 (January): 5–15.

MEYER, LAURENCE H. 1998. "Start with a Paradigm, End with a Story: The Value of Model-Based Forecasting." Speech to the Downtown Economics Club, New York, New York, June 3.

PAGAN, ADRIAN, AND JOHN ROBERTSON. 1995. "Resolving the Liquidity Effect." Federal Reserve Bank of St. Louis *Review* 77 (May–June): 33–54.

ROBERT, CHRISTIAN P. 1994. *The Bayesian Choice: A Decision-Theoretic Motivation*. New York: Springer-Verlag.

ROBERTSON, JOHN C., AND ELLIS W. TALLMAN. 1999a. "Improving Forecasts of the Federal Funds Rate in a Policy Model." Federal Reserve Bank of Atlanta Working Paper 99-3, March.

———. 1999b. "Vector Autoregressions: Forecasting and Reality." Federal Reserve Bank of Atlanta *Economic Review* 84 (First Quarter): 4–18.

RUDEBUSCH, GLENN D. 1998. "Do Measures of Monetary Policy in A VAR Make Sense?" *International Economic Review* 39 (November): 907–31.

SHAPIRO, MATTHEW. 1994. "Federal Reserve Policy: Cause and Effect." In *Monetary Policy*, edited by N. Gregory Mankiw. Chicago: University of Chicago Press.

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

———. 1986. "Are Forecasting Models Usable for Policy Analysis?" Federal Reserve Bank of Minneapolis *Quarterly Review* (Winter): 2–16.

———. 1993. "A Nine Variable Probabilistic Macroeconomic Forecasting Model." In *Business Cycles, Indicators, and Forecasting*, edited by James H. Stock and Mark W. Watson. NBER Studies in Business Cycles, vol. 28. Chicago: University of Chicago Press.

———. 1996. "Macroeconomics and Methodology." *Journal of Economic Perspectives* 10 (Winter): 105–20.

SIMS, CHRISTOPHER A., AND TAO ZHA. 1998. "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review* 39 (November): 949–68.

STAIGER, DOUGLAS, JAMES STOCK, AND MARK WATSON. 1997. "The NAIRU, Unemployment, and Monetary Policy." *Journal of Economic Perspectives* 11 (Winter): 33–49.

STRONGIN, STEVEN. 1995. "The Identification of Monetary Policy Disturbances: Explaining the Liquidity Puzzle." *Journal of Monetary Economics* 35 (August): 463–97.

TOBIN, JAMES. 1970. "Money and Income: *Post Hoc Ergo Propter Hoc*?" *Quarterly Journal of Economics* 84 (May): 301–17.

UHLIG, HARALD. 1997. "Bayesian Vector Autoregressions with Stochastic Volatility." *Econometrica* 65 (January): 59–73.

WAGGONER, DANIEL F., AND TAO ZHA. 1998. "Conditional Forecasts in Dynamic Multivariate Models." Federal Reserve Bank of Atlanta Working Paper 98-22, December.

———. 1999. "Does Normalization Matter for Inference?" Federal Reserve Bank of Atlanta. Unpublished paper.

ZHA, TAO. 1997. "Identifying Monetary Policy: A Primer." Federal Reserve Bank of Atlanta *Economic Review* 82 (Second Quarter): 26–43.

———. 1998. "A Dynamic Multivariate Model for Use in Formulating Policy." Federal Reserve Bank of Atlanta *Economic Review* 83 (First Quarter): 16–29.

Fully Funded Social Security: Now You See It, Now You Don't?

MARCO A. ESPINOSA-VEGA AND STEVEN RUSSELL
Espinosa-Vega is a senior economist in the Atlanta Fed's research department. Russell is an associate professor of economics at Indiana University–Purdue University Indianapolis. *They thank Zsolt Becsi, Alex Mourmouras, Peter Rangazas, and Tapen Sinha for their very useful comments. The authors are solely responsible for content.*

A SOCIAL SECURITY SYSTEM IS A SYSTEM THAT USES REVENUE FROM TAXES ON WAGE INCOME TO PROVIDE PAYMENTS TO SENIOR CITIZENS. ONE OF THE MAIN GOALS OF SUCH A SYSTEM IS TO ENSURE THAT ELDERLY PEOPLE HAVE ADEQUATE INCOMES WHEN THEY RETIRE.[1] MOST DEVELOPED COUNTRIES ESTABLISHED SOCIAL SECURITY SYSTEMS AT SOME POINT DURING THE EARLY PART OF THIS CENTURY. THE UNITED STATES ESTABLISHED ITS SOCIAL SECURITY SYSTEM DURING THE MID-1930S, WHEN WIDESPREAD POVERTY PRODUCED BY THE GREAT DEPRESSION INCREASED PUBLIC CONCERN FOR THE PLIGHT OF ELDERLY PEOPLE AND OTHERS WITH SPECIAL NEEDS.[2] MEXICO ESTABLISHED ITS SOCIAL SECURITY SYSTEM LESS THAN A DECADE LATER, AS PART OF THE SOCIAL REFORM PROGRAM BEGUN BY THE REVERED LEADER LAZARO CARDENAS. CARDENAS WAS ONE OF THE FOUNDERS OF THE INSTITUTIONAL REVOLUTIONARY PARTY, WHICH HAS GOVERNED MEXICO FOR THE LAST SIXTY YEARS.

Mexico's social security system is of special interest to economists in the United States. Part of the reason for this interest, of course, is that Mexico is a neighboring country with which the United States has a close social, political, and economic relationship. The primary reason, however, is that the government of Mexico has recently implemented a social security reform program that is broadly similar to the social security reform programs advocated by many U.S. economists and policymakers.

Social security systems can be implemented in a number of different ways. These differences can have important effects on macroeconomic variables such as national saving, interest rates, investment, and growth. One particularly key distinction between different types of social security systems is the distinction between "pay-as-you-go" systems and "fully funded" systems. The principal goals of this article are to describe the basic differences between these two types of social security system, to indicate why these differences are important,

and to explain, using Mexico as an example, why it may be surprisingly difficult to determine which type of system a country actually has.

In any social security system, current workers pay taxes that are called their contributions to the system and retired workers receive payments, called their benefits, from the system. Under a pay-as-you-go system, the contributions of current workers are used, directly or indirectly, to pay benefits to current retirees. Under a fully funded system, in contrast, the contributions of current workers are not used to pay current benefits. Instead, these contributions are used to purchase assets, and the returns on these assets—that is, their principal and interest—are used to pay the future social security benefits of the workers who made the contributions. Thus, under a fully funded system there is a sense in which retired workers have financed their own benefits while under a pay-as-you-go system the benefits of retired workers are financed by current workers.

Although it is fairly easy to explain the theoretical difference between the two kinds of social security systems, as a practical matter it may not be so easy to determine which type of system a particular country has. The current situation in Mexico provides an interesting and instructive example of this practical problem. According to the Mexican government, one of the principal goals of its social security reform program is to convert the country's pay-as-you-go social security system into a fully funded system. As this article will explain, however, there may be some room for doubt that Mexico's new social security system is or ever will be fully funded. Instead, the new system may be a pay-as-you-go system of a somewhat different type. While a switch of this type may have some economic benefits, these benefits are likely to be considerably smaller than the ones produced by a genuine switch to a fully funded social security system.

It is important to note that Mexico is reforming its social security system in a number of different ways, many of which are not directly relevant to the question of whether the new system is fully funded. Some aspects of the reform program may represent significant improvements over the old social security system and may produce substantial benefits for the Mexican public, even if the new system does not turn out to be fully funded. A companion article scheduled for publication in the next issue of this *Economic Review* presents a detailed description of Mexico's old and new social security systems. This article confines itself to (1) presenting a general discussion of the characteristics of pay-as-you-go versus fully funded social security systems and (2) identifying the issues involved in determining whether Mexico's new social security system is fully funded.

Pay-As-You-Go Systems

Pay-as-you-go social security systems also come in two basic types. The simpler type is one in which the government social security agency—in the United States, the Social Security Administration—collects taxes from workers and uses them to make direct payments, sometimes called transfers, to retired people. In this article, a system of this type is referred to as a tax/transfer system.

Social security systems can be implemented in a number of different ways. These differences can have important effects on macroeconomic variables such as national saving, interest rates, investment, and growth.

In an alternative type of pay-as-you-go social security system, the government social security agency uses workers' contributions to purchase financial assets. Typically, these assets are bonds issued by a government budget agency, such as the U.S. Treasury Department, although they may also include private bonds or stocks (see below). For the moment, however, it is simplest to assume that the assets consist exclusively of government bonds. In this case, the government budget agency uses the proceeds of the sale of new bonds to pay off previously issued bonds. These bond-financed repayments constitute the social security benefits of current retirees: the social security system bought the maturing bonds using their past contributions. The principal and interest on the currently issued bonds will constitute the social security benefits of future retirees. The government budget agency will pay these bond returns by issuing new bonds to the social security agency, the agency will buy them using the contributions of future workers, and so on. In this article, a pay-as-you-go

1. *Social security systems may also make payments to people who are not retired, such as workers who become disabled or dependents of deceased workers.*
2. *Sargent (1998) points out that the U.S. social security system was established at a time when academic economists believed excessive saving and overaccumulation of capital were a significant problem. In his view, this belief played a role in building support for the system.*

system of this type will be referred to as a bond/tax-or-transfer system.[3]

In a bond/tax-or-transfer system, if the social security agency wishes to pay benefits that are larger than the bond repayments, then it will have to ask the government budget agency for funds it can use to make supplemental transfers to retirees. The budget agency obtains these funds by selling more new bonds each year than it needs to obtain the funds necessary to pay off its maturing bonds. If the social security agency plans to pay benefits that are smaller than the bond returns, then it can ask the government to levy taxes on retired people that are equal to the difference between the bond returns and the desired benefits. The budget agency can use this tax revenue to reduce the quantity of new bonds it needs to issue to finance current social security benefits.

Fully funded social security systems are profoundly different from pay-as-you-go systems, and a successful transition to a fully funded system might have very significant long-run macroeconomic benefits for a national economy.

There is no fundamental difference between tax/transfer and bond/tax-or-transfer systems.[4] Under a tax/transfer system, the government social security agency collects contributions, in the form of taxes, from workers while they are working. It gives these contributions to retired people as their benefits. The agency also promises workers transfer benefits when they retire. It will get these benefits by taxing future workers, and so on. Under a bond/tax-or-transfer system, the social security agency also collects contributions from workers while they are working. It uses the contributions to buy newly issued bonds from the budget agency. However, the budget agency will use the contributions to pay off the bonds it sold to the social security agency a generation ago—when the currently retired workers were still working—and the social security agency will use these dollars to pay these workers' current social security benefits. Thus, current social security contributions end up in the hands of current retirees as benefits, just as under the tax/transfer system. Similarly, although workers will receive future social security benefits that are based, for accounting purposes, on the returns on bonds that the social security agency purchased with their contributions, these returns will actually be financed by the sale of new bonds to the agency. The funds to purchase these bonds will be provided by the social security contributions of young people who will be working when older workers are retired. In a practical sense, the older workers' benefits will come from the contributions of these young workers.

To understand the equivalence between these two different varieties of a pay-as-you-go system, it is important to remember that a government bond is simply a promise by the government to make a payment in the future. A government promise to make a payment to pay off a bond is not fundamentally different from a government promise to make a payment for social security benefits. If the government requires workers to buy bonds and promises them future payments to retire the bonds, then it is not doing anything essentially different from requiring them to pay taxes and promising them a future transfer payment.

As indicated above, the government of Mexico recently began implementing a social security reform program that it describes as involving a switch from a pay-as-you-go system to a fully funded system. This article argues that it is quite possible that the reform program represents a change of a much less substantive sort: a switch from a pay-as-you-go system of the tax/transfer type to a pay-as-you-go system of the bond/tax-or-transfer type.

Distinguishing Pay-As-You-Go from Fully Funded Systems

One source of confusion in distinguishing pay-as-you-go systems from fully funded systems is the fact that it is possible, under either system, for social security contributions to be used to purchase financial assets that include government bonds. As the previous section explained, under a pay-as-you-go system of the bond/tax-or-transfer type the social security contributions are used to purchase financial assets, and the assets in question may be government bonds. Under a fully funded system the social security contributions are always used to purchase financial assets; again, these assets may be government bonds.

When both types of systems purchase government bonds, the important distinction between them involves the question of why the government bonds are being issued—that is, what the proceeds of the bond sales are being used for. Under a pay-as-you-go system, when the social security system uses the contributions of current workers to purchase newly issued government bonds, the government uses the proceeds of the bond sales to pay off old bonds that were issued to finance social security payments to past workers. The existence of the social security system provides the only reason for the government to issue the new bonds, and it also provided the only reason the government needed to issue the old bonds.

Under a fully funded system, in contrast, the government bonds that the social security system purchases were issued for some other purpose—for instance,

to finance a current government project or to roll over bonds originally issued to finance a past project. The government does not use the proceeds of these bond sales to retire bonds that were issued to pay social security benefits, and the bonds would have been issued even in the absence of a social security system.

Why is the question of how the government uses the bonds it sells to the social security system so important? As the introduction notes, the key feature that distinguishes a pay-as-you-go system from a fully funded system is the source of the funds used to pay benefits to retired workers: do these funds come from current workers, or do they come from the returns on the retired workers' assets? If the funds used to pay benefits to retired workers are obtained by selling bonds to current workers, then it is clear that the current workers are the source of the retired workers' benefits. In this case, the system is not fully funded: the retired workers have not financed their own benefits, and current workers will not have to increase their saving to finance their future benefits (see below).

Defined Contributions versus Defined Benefits

Another important feature that distinguishes some social security systems from others is the nature of the relationship between the size of a worker's current social security contributions and the size of the same worker's future social security benefits. Under a defined contributions system, a worker's social security contributions are used to purchase assets, and the size of the worker's benefits depends on the rate of return on those assets. If the rate of return on the assets turns out to be high, then the worker will receive relatively large retirement benefits, and vice versa.[5] Under a defined benefits system, in contrast, the social security benefits paid to a retired worker are determined by a fixed formula that involves factors like total contributions to the system, total number of years worked, salary during the last few years before retirement, age at retirement, and so on. Workers' social security contributions may be used to purchase assets or to finance direct transfers to retirees, but in either case the workers' retirement benefits do not depend on the returns on any assets.

Currently, the United States has a defined benefits system. Before its recent social security reform, Mexico also had a defined benefits system. Mexico's new system features defined contributions. From the point of view of workers, the attraction of a defined benefits system is that it reduces the amount of uncertainty about the value of their future benefits. On the other hand, a defined benefits system produces considerable uncertainty for the government, which usually finances the future benefits out of revenue from taxes or assets whose value depends on future economic conditions. If the promised benefits turn out to be larger than the amount of revenue, then the government has to obtain supplementary financing by borrowing or by increasing taxes.

In the United States, Mexico, and many other countries, demographic changes are producing a rapid increase in the fraction of the population that consists of retired workers. As a result, the value of the social security contributions from young workers is growing more slowly than the value of the defined benefits due old workers. This situation has produced serious financial stresses. It is a big part of the reason that many countries have switched or are considering switching to defined contributions systems.

Under a fully funded system there is a sense in which retired workers have financed their own benefits while under a pay-as-you-go system the benefits of retired workers are financed by current workers.

Historically, pay-as-you-go social security systems have usually featured defined benefits, and fully funded systems have usually featured defined contributions. Other combinations are possible, however. Under a pay-as-you-go system with defined contributions, the benefits associated with a worker's current contributions could be indexed to the value of the future revenue produced by a fixed social security tax rate, allowing the level of benefits to vary with the economy's demographic evolution and growth performance. Under a fully funded system with defined benefits, the government could specify workers' future benefits and cover

3. *The U.S. social security system includes a social security trust fund that holds U.S. Treasury bonds. However, most social security contributions are used more or less immediately to finance social security benefits. The social security trust fund acts as a buffer to help smooth out temporary differences between total current contributions and total current benefits. If current contributions are larger than current benefits, then the surplus contributions are used to purchase bonds. In the opposite case, some of the bonds are sold in order to provide a supplementary source of funds for the benefits.*
4. *This point is made briefly in Auerbach and Kotlikoff (1998, 161–62) and more completely in Murphy and Welch (1998).*
5. *In many cases, workers are not allowed to withdraw the entire value of their assets as a lump sum. Instead, the funds must be used to buy an annuity. In some cases, workers are allowed to withdraw their funds according to a schedule that permits them to withdraw a fixed amount each year over a fixed number of years.*

any asset-returns shortfall via taxes or borrowing. Moreover, if the assets held by the system consist mostly of government debt, as in the case of Mexico, then there should be little difference between defined contributions and defined benefits. Presumably, the government's promise to repay its debt is no more or less reliable than its promise to pay future social security benefits directly.[6]

Privatized Social Security Systems

Another source of confusion about different types of social security systems involves the concept of a privatized social security system. The confusion occurs because the term privatized can be used in connection with several different aspects of social security systems. Sometimes it refers to who manages the system, sometimes to the relationship between a worker's current social security contributions and the same worker's future benefits, and sometimes to the type of assets held by the social security system.

According to the Mexican government, one of the principal goals of its social security reform program is to convert the country's pay-as-you-go social security system into a fully funded system.

Management versus Financing. One usage of *privatized* in the context of social security refers to the degree of government involvement in managing the system. If a social security system is extensively privatized in this sense, then the system is managed mostly by private firms, leaving a relatively limited role for the government. Under a conventional or unprivatized system, on the other hand, the government plays an exclusive or leading role in managing the system.

A second usage of *privatized* involves the ultimate source of the funds used to pay social security benefits. As noted above, under a fully funded social security system the social security benefits paid to a retired worker are financed by the worker's contributions made before retirement. This situation is often summarized by saying that the worker's benefits have been financed privately. Under a pay-as-you-go system, on the other hand, a worker's benefits are ultimately financed by the contributions of other workers, so they could be described as having been financed publicly rather than privately.

When a government announces that it is planning to privatize a social security system, people often assume that the government is planning to switch from a pay-as-you-go system to a fully funded system. However, the government may simply intend to turn the task of administering the system over to private firms. A pay-as-you-go social security system can be administered by the private sector or by the government; similarly, a fully funded system can be administered by the private sector or by the government.

To understand how a pay-as-you-go social security system can be privately administered, imagine a bond/tax-or-transfer system in which the government allows private financial intermediaries to set up social security accounts on behalf of workers. The intermediaries would purchase and sell bonds, make social security benefit payments, and provide associated accounting services. Workers might be allowed to choose which private intermediary would receive their funds and pay their benefits. The intermediaries might have considerable latitude about which government securities to buy, what sorts of accounting systems to use, what handling fees to charge, and so forth. This is the type of system that may have been established in Mexico.

To see how a fully funded system can be government administered, imagine a government social security agency that maintained social security accounts on behalf of active and retired workers. The agency would decide which government or private securities to purchase. It would handle all the accounting, and it would pay all the benefits. The benefits, however, would be funded entirely from the principal and interest on these securities, and the securities would have been issued for purposes unrelated to the needs of the social security system.

In order to avoid confusion, it might be better if the term privatized were used only to describe fully funded social security systems. It is difficult, however, for people to think of a social security system that is mostly privately administered as anything other than privatized, even though the system may be of the pay-as-you-go type.

Asset Holdings. Finally, a social security system might be described as privatized if the assets the system purchases are privately issued—that is, if they consist of corporate bonds or stock rather than government bonds. As noted earlier, people tend to think of fully funded systems as privatized and of pay-as-you-go systems as dominated by the government. Consequently, they often assume that it is possible to distinguish between pay-as-you-go systems and fully funded systems on the basis of whether the systems' assets are issued by the private sector or by the government. They may also assume that the economic impact of a given type of system depends on the type of assets it holds.

In reality, a pay-as-you-go social security system may hold privately issued or government-issued assets, and the same is true for a fully funded system. In addition, the economic impact of a social security system may not depend on the type of assets it holds. It is pos-

sible, for example, for a fully funded system that holds privately issued assets to have the same economic impact as a fully funded system that holds only government debt. Moreover, a system that holds privately issued assets and appears fully funded to casual observers may have exactly the same economic impact as a pay-as-you-go system.

Macroeconomic Implications

One of the principal goals of this article is to analyze the economic implications of alternative types of social security systems. To accomplish this goal, it is important to consider both the asset portfolio held by the social security system on behalf of current and past contributors and the overall asset portfolio of the public and the government.

Portfolio Substitution and Its Implications. Consider, for example, the case of a pay-as-you-go system of the bond/tax-or-transfer type that is being established in an economy that did not previously have a social security system. In Case A, the social security system holds only government bonds. In Case B, the system is allowed to hold private bonds and for the most part chooses to do so. (As a result, it may be mistaken for a fully funded system.)

Suppose that in both cases the government initiates the new system by issuing bonds to finance the social security payments made to current or near-future retirees. In Case A, these bonds will be purchased by the social security system. In Case B, on the other hand, the social security system does not purchase many government bonds. Does this mean that the government will be unable to market its initial bond issue so that the new social security system will collapse as it begins? Does it mean that the government will be able to sell its bonds only by offering very high interest rates that will drive up other market rates and have adverse repercussions across the economy?

Probably not. Consider Case A. Although the new social security system is likely to reduce the amount of private saving (as seen below), there probably will still be a significant amount of private saving. Much of this private saving is likely to take the form of purchases of private bonds. In Case B, there will also be substantial private saving, but many of the private bonds that workers would have liked to purchase will be purchased by the social security system. This situation will cause a decrease in the supply of private bonds. But since government bonds and private bonds are likely to be good substitutes, active workers who would have saved by buying private bonds (Case A) will now save by buying government bonds (Case B). Thus, the government will have no trouble selling its bonds at moderate interest rates, even though the social security system may not be buying them. Although the workers' social security benefits will now be based on the returns from private bonds and their private retirement income will be based on returns from government bonds, their total retirement income will be unchanged. Moreover, the funds the government uses to pay benefits to currently retired workers will continue to come from currently active workers, just as under any pay-as-you-go system.

Similarly, suppose (Case C) that a new social security system is allowed to hold stocks and for the most part chooses to do so. Stock portfolios have higher average returns than portfolios of government or private bonds, but they are also riskier. Thus, in Case C the benefits paid by the social security system will be higher, on average, than the benefits paid by the system in Case A, but their value will also be riskier. (This discussion assumes that the system features defined contributions.) Returning again to Case A, in which the social security system purchases only government bonds, the private saving that occurs outside the social security system is likely to include a substantial quantity of stocks. Workers will purchase these stocks because they like the high average returns and are willing to accept the increased risk, up to a point.

In Case C, however, workers will get high average returns from the social security system but also high risk. They are likely to want less risk in their private saving, and they are likely to be willing to accept lower returns in order to obtain lower risk. As a result, they will reduce their private stock purchases by the amount of stock the social security system purchases on their behalf. They will replace these stocks with safer government bonds. Thus, neither the total amount of retirement-income risk the workers will be accepting nor the total amount of income they will receive when they retire will be different from that in Case A. Again, the government will be able to sell its bonds at moderate interest rates even though the social security system will not be buying them. Again, the funds used to pay current social security benefits ultimately come from currently active workers.

Under a pay-as-you-go bond/tax-or-transfer system that features defined benefits, on the other hand, a social security reform program under which the system switches from holding bonds to holding stocks may have significant economic effects. Before the reform, the market presumably determined stock prices and return rates in a way that ensured that all the stock issued by

6. This discussion assumes that the social security system has the option of purchasing government bonds that are indexed against inflation. The new Mexican social security system has large holdings of indexed government bonds.

firms would be purchased by active workers as part of their private saving. When these workers reached retirement age, they would end up with relatively high retirement income if the stocks performed well but relatively low income if the stocks performed badly. After the reform, however, the social security system will be purchasing stocks. The market will have to adjust stock and bond prices and return rates in a way that induces young workers to increase their purchases of less risky assets (government or private bonds) even though their social security retirement income is not risky. If the rate of return on the stocks turns out to be relatively low, then the government may have to supplement the stock returns by increasing taxes on the active workers. If the return rate on the stocks turns out to be relatively high, then the government may be able to reduce taxes. The upshot is that allowing the social security system to hold stocks shifts risk from workers' retirement years to the years when they are still working. Retired workers receive income that is much more predictable than it otherwise would have been, but some generations of active workers end up paying relatively high taxes while others end up paying relatively low taxes.

There may be some room for doubt that Mexico's new social security system is or ever will be fully funded. Instead, the new system may be a pay-as-you-go system of a somewhat different type.

Saving and Interest Rates. As the introduction indicated, the principal reason that the distinction between pay-as-you-go social security systems and fully funded systems is potentially important is that the two different types of systems may have very different effects on saving, interest rates, and related macroeconomic variables. Perhaps the easiest way to see how and why the macroeconomic effects of the two systems differ is by conducting the "thought experiment" of starting with an economy that has no social security system and then introducing such a system. The effects of introducing a pay-as-you-go social security system will turn out to be very different from the effects of introducing a fully funded system.[7]

In the economy without a social security system, people will have to save substantial amounts while they are working in order to provide funds to support themselves after they retire. Under a social security system, regardless of which type, by contrast, active workers will be paying taxes they would not be paying otherwise. They also know they will have substantial retirement income even if they do not save large amounts. As a result, they are likely to reduce their current saving in order to try to restore their current consumption to its original level.

The reduction in current saving by workers (current private saving) will reduce the availability of credit in the economy. Under a pay-as-you-go system, moreover, the government will use the social security tax revenue to pay social security benefits (directly or indirectly, as discussed above), so it will not be able to use this revenue to reduce its borrowing. As a result, there will be no change in the amount of credit the economy requires. In terms of conventional supply-demand analysis, the credit supply curve will shift to the left along an unchanged credit demand curve. As a result, the equilibrium quantities of saving and credit will fall and the equilibrium interest rate will rise. Thus, a basic prediction of social security theory is that establishing a pay-as-you-go system should cause the amount of saving in an economy to fall and the interest rate in the economy to rise.

Next, imagine introducing a fully funded social security system into an economy that has not had a social security system. Under a fully funded system, the combination of current social security taxes and expected future social security benefits again leads to a decrease in saving by workers, reducing the availability of credit. In this case, however, the government does not use the social security tax revenue to pay current social security benefits; instead, it uses the revenue to retire existing government bonds. The retirement of these bonds will reduce the government's debt service payments in the future, creating surplus funds that it can use to pay future social security benefits without borrowing.

In terms of supply-demand analysis, under the fully funded system the leftward shift in the private supply of credit is accompanied by a leftward shift in government demand for credit of roughly equal size. As a result, there should be no significant change in interest rates. Private saving has decreased, but since government dissaving (borrowing) has decreased by a roughly equal amount, there should be no significant change in total saving in the economy.

Thus, a second basic prediction of social security theory is that establishing a fully funded social security system should have little or no effect on the economy. Stated differently, an economy with a fully funded social security system is not much different from an economy with no social security system.[8] It follows that switching from a pay-as-you-go social security system to a fully funded system has roughly the same economic impact as eliminating a pay-as-you-go system without replacing it. Consequently, switching from a pay-as-you-go system to a fully funded system should cause the total amount of saving to rise, producing a decline in the interest rate.[9]

Public Welfare. An important follow-up question is whether an increase in the amount of saving is good or bad for the economy. Since the principal role of saving in an economy is to finance the acquisition of physical capital, this question becomes the question of whether the economy would be better off trying to maintain a larger stock of capital. According to economic theory, the answer to this question depends on whether the long-run return rate on capital is higher or lower than the long-run growth rate of the economy—in the jargon of economic theorists, whether the economy is dynamically efficient or dynamically inefficient.[10] If the rate of return on capital is relatively high, so that the economy is dynamically efficient, then capital is productive at the margin. The next units of capital acquired via saving will produce additional goods in the future whose value exceeds the amount of future saving and investment that will be required to maintain them. In this case, an increase in the amount of saving makes the economy better off in the long run, and vice versa. Consequently, switching from a pay-as-you-go system to a fully funded system would make the economy better off in the long run.

In the opposite case, where the rate of return on capital is relatively low, so that the economy is dynamically inefficient, the last units of capital are unproductive at the margin. The amount of saving and investment necessary to maintain these units of capital is larger than the amount of additional goods they will produce. As a result, the economy could actually consume more each year by saving less and reducing both its annual investment and its capital stock. In this case, switching from a pay-as-you-go system to a fully funded system would make the economy worse off in the long run because the amount of saving would rise and the degree of dynamic inefficiency would increase.

Although there is some debate about the issue, most economists believe that most modern economies, certainly the U.S. economy and presumably the Mexican economy, are dynamically efficient.[11] As a result, most economists believe that switching to fully funded social security systems would make these economies better off in the long run. It seems likely that this belief is a large and perhaps dominant part of the reason the Mexican government would like to switch to a fully funded system.

Transition Problems

Unfortunately for the Mexican government, and for other governments interested in engineering this kind of switch, the price of achieving the long-run gain from switching systems may be considerable pain in the short run. For the reasons just outlined, switching to a fully funded social security system is arguably likely to benefit workers born in the relatively distant future. However, it is almost certain to hurt many current workers, active or retired, and it may also hurt workers who are born or who retire in the near future.

The biggest problem in managing the transition from a pay-as-you-go system to a fully funded system is how to finance the benefits that were due under the old system to workers who have already retired or who will retire in the near future. Under the old pay-as-you-go system, these benefits were to have been financed out of the social security contributions of current workers. Under the new fully funded system, however, current social security contributions must be used to purchase assets, and the sellers of these assets must use the funds for some sort of investment rather than transferring them to current retired workers. Consequently, an immediate switch to a fully funded system would deprive current and near-future retirees of their social security benefits, leaving many of them with little or no retirement income. Since current and near-future retirees are also current voters, it is likely that the government of a democratic or quasidemocratic country would face serious political opposition to trying to carry out such a switch. In the words of Thomas Sargent, "it is easier to vote an unfunded social retirement system in than to vote one out"(1998, 306).[12]

7. The discussion of the macroeconomic effects of social security systems will assume that workers' economic decisions are not very strongly influenced by altruistic feelings toward their ancestors or descendants. Most economists believe this assumption is appropriate, at least as a first approximation. Broadly speaking, the presence of intergenerational altruism tends to reduce the difference between the macroeconomic effects of pay-as-you-go versus fully funded systems.

8. Making this statement ignores a number of possible microeconomic effects of the establishment of social security systems—in particular, the tendency of many systems to redistribute income toward low-income people by giving them relatively generous benefits. It also ignores the possibility that some people are shortsighted and will not save economically rational amounts unless a social security system forces them to do so. Both these considerations have figured prominently in practical discussions of social security reform in the United States, Mexico, and elsewhere.

9. This point has been emphasized by Kotlikoff (1998), among many others.

10. One of the first economists to recognize the possibility of dynamic inefficiency was Samuelson (1958).

11. Abel and others (1989) make the case that the United States and a number of other developed economies are dynamically efficient.

12. Cooley and Soares (1999) discuss the possibility that a pay-as-you-go social security system may represent the outcome of a democratic political process in which different groups support policies that reflect their own economic interests.

An alternative method for executing an immediate switch to a fully funded system would be to finance the social security benefits due current and near-future retirees by some combination of increased taxes and cuts in government expenditures—including, possibly, reductions in the generosity of the social security benefits. Again, however, policies of this sort would impose a large financial burden on current workers and other groups of potential voters. In Mexico's case, moreover, a financing policy of this sort seems doubly unlikely because the government is trying to reform its social security system in the aftermath of an economic crisis that has sharply reduced workers' incomes and living standards.

Some aspects of Mexico's reform program may represent significant improvements over the old social security system and may produce substantial benefits for the Mexican public even if the new system does not turn out to be fully funded.

The transition strategies that seem most likely to be politically feasible would involve spreading the burden of financing the social security benefits due current and near-future retirees across a number of future generations of workers. Under a strategy of this type, the Mexican government would start by issuing long-term bonds in order to obtain the funds necessary to pay social security benefits to current and near-future retirees. When these bonds mature, the current workers will be retired, and they will have been replaced by a new generation of workers. At this point, the government would increase the taxes on current workers in order to obtain the funds needed to retire some of the bonds. The remainder of the bonds would be rolled over. When the second round of bonds matured, the government would use the same supplementary tax revenue—now collected from a second new generation of workers—to retire some additional bonds; it would roll over the rest, and so on. Eventually, there would be no bonds left to roll over, so the original debt would be fully retired. The government would have completed the transition from a pay-as-you-go social security system to a fully funded system.

It is now possible to pose the key question that provided the motivation for writing this article. How can an analyst observing the actions of a government that is implementing a social security reform program—a program which, according to the government, will convert the country's social security system from pay-as-you-go to fully funded—determine whether the government is really switching to a fully funded social security system as opposed to simply changing the form of the pay-as-you-go system?

As the discussion presented earlier in this section indicates, a central question in trying to determine the nature of a social security reform program is how the establishment of the program affects the government's overall budget deficit. If a government that is trying to switch to a fully funded social security system manages to pay the current social security benefits without increasing its budget deficit, then it may have financed these benefits via tax increases, spending cuts, or some combination of the two. In this case, it may have succeeded in engineering an immediate transition. On the other hand, if the government deficit rises by an amount equal to the total cost of paying the current social security benefits, then the government has presumably financed these benefits by additional borrowing, which means it has not yet taken any firm steps toward a successful transition. In the intermediate case, in which the government deficit increases by an amount that is smaller than the cost of the current social security benefits, the size of the step taken toward a successful transition can be measured by the fraction of the current benefits that is not covered by an increase in the deficit. [13]

Suppose that the fraction of the current social security benefits that the government is able to finance by spending cuts or tax increases is relatively small. In this case, how is it possible to tell whether there is likely to be a genuine transition to a fully funded system? The answer, it turns out, is "not very easily." The basic reason for the uncertainty is that the actions the government must take at the beginning of the transition process—the only actions our imaginary analyst can observe—are exactly the same in both cases: it must issue bonds to obtain the funds needed to pay most of the social security payments due current and near-future retirees. The government actions that will distinguish a transition to a fully funded system from a transition to a pay-as-you-go bond/tax-or-transfer system will occur in the future, not today. If the government is really switching to a fully funded system, then over the next few generations it will have to collect enough additional revenue, via new taxes or cuts in spending, to retire the aforementioned bonds. But if it is simply switching to a pay-as-you-go system of the bond/tax-or-transfer type, then it will not have to reduce its future budget deficits because it will roll the bonds over indefinitely without retiring any of them.[14]

Although a switch of the latter sort may have few or no economic effects, it creates the appearance of reform in two different ways. First, since switching to a bond-based system could (but does not necessarily) represent the first step in a transition to a fully funded system, this switch allows the government to claim that it has begun the transition process. Second, the switch

to a bond-based system allows the government to privatize a number of aspects of the administration of the social security system. This step may have some benefits in its own right, and many people are likely to misinterpret it as representing more effectual reform.

Conclusion

Governments of countries around the world, including Mexico and the United States, have implemented or are considering implementing social security reform programs. In virtually every case, one of the principal goals of the reform program is to convert a pay-as-you-go social security system into a fully funded system.

Fully funded social security systems are profoundly different from pay-as-you-go systems, and a successful transition to a fully funded system might have very significant long-run macroeconomic benefits for a national economy. However, it is not always easy to determine whether a country has a pay-as-you-go system or a fully funded system, and it may be even more difficult to determine whether a country is likely to succeed in switching from one type of system to the other. The economic circumstances of most of the countries that are conducting or contemplating social security reforms will force them to proceed with these reforms in a very gradual way. Many of the steps a country might take in order to begin a gradual transition to a fully funded social security system are identical to steps it might take if it is merely changing the form of its pay-as-you-go system—a change whose long-run macroeconomic benefits may not be very significant. And since the actions needed to push the transition process forward may have substantial political costs, governments have a potential incentive to claim that they intend to make the switch even when they have no such intention. Even if the government genuinely intends and expects a transition to take place, carrying out the transition will require cooperation from future governments, and these governments will also have powerful incentives not to take the steps needed. Finally, even if the public believes that a successful transition will occur, the fact that the transition is likely to be quite gradual means that the changes in their current behavior resulting from this belief may not be large enough to be identified with any confidence.

The bottom line is that information that is currently available, or that will become available in the near future, may give very little indication as to whether Mexico or other countries attempting gradual reforms are likely to succeed in replacing their pay-as-you-go social security systems with fully funded systems. The information needed to make this judgment is likely to be revealed very slowly over time.

13. In practice, unfortunately, making judgments like this can be quite challenging. Interpreting government budget statistics is often very difficult, and the budget of the social security system is often reported separately from the rest of the government budget.

14. The government will have to pay the interest on the bonds, but it can do so without increasing its social security tax collections.

REFERENCES

ABEL, ANDREW, GREGORY N. MANKIW, LAWRENCE H. SUMMERS, AND RICHARD J. ZECKHAUSER. 1989. "Assessing Dynamic Efficiency: Theory and Evidence." *Review of Economic Studies* 56 (January): 1–19.

AUERBACH, ALAN J., AND LAURENCE J. KOTLIKOFF. 1998. *Macroeconomics: An Integrated Approach*. Cambridge, Mass.: MIT Press.

COOLEY, THOMAS, AND JORGE SOARES. 1999. "Social Security Based on Reputation." *Journal of Political Economy* 107 (February): 135–60.

KOTLIKOFF, LAURENCE J. 1998. "Simulating the Privatization of Social Security in General Equilibrium." In *Privatizing Social Security*, edited by Martin Feldstein. Chicago: University of Chicago Press.

MURPHY, KEVIN M., AND FINIS WELCH. 1998. "Perspectives on the Social Security Crisis and Proposed Solutions." *American Economic Review* 88 (May): 142–50.

SAMUELSON, PAUL A. 1958. "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money." *Journal of Political Economy* 66 (December): 467–82.

SARGENT, THOMAS. 1998. "Comment on 'Stimulating the Privatization of Social Security.' " In *Privatizing Social Security*, edited by Martin Feldstein. Chicago: University of Chicago Press.

Mergers of Publicly Traded Banking Organizations Revisited

SIMON KWAN AND ROBERT A. EISENBEIS
Kwan is a financial economist at the Federal Reserve Bank of San Francisco. Eisenbeis is a senior vice president and the director of research at the Federal Reserve Bank of Atlanta. *The authors acknowledge the helpful comments of Frank King and Steven J. Pilloff.*

THE PAST FEW YEARS HAVE SEEN A SURGE IN BANKING MERGERS. IN MORE THAN 3,844 MERGERS AND ACQUISITIONS BETWEEN 1989 AND 1999, ACQUIRING INSTITUTIONS PURCHASED MORE THAN $3 TRILLION IN ASSETS (SEE TABLE 1). A NUMBER OF REASONS HAVE BEEN ADVANCED FOR SUCH ACQUISITIONS, INCLUDING THE NEED TO CONSOLIDATE TO ACHIEVE COST SAVINGS AND OPERATIONAL EFFICIENCIES, TO BE BETTER ABLE TO COMPETE IN THE GLOBAL MARKETPLACE, OR TO PROVIDE FOR THE CONTROLLED EXIT OF INEFFICIENT FIRMS FROM THE FINANCIAL SERVICES INDUSTRY.

This article adds to the limited existing research on the effects of bank mergers by analyzing consolidations between 1989 and 1996, a period of almost unprecedented banking consolidation involving acquisitions impossible in earlier times. Consistent with previous studies, the findings suggest that the various expected performance and earnings benefits of mergers may not in fact be realized.

Earlier Studies

Evidence supporting consolidation to achieve cost savings and operational efficiencies is sparse. Pilloff and Santomero (1997) review the research evidence and claims for efficiency gains, concluding that there is little empirical evidence of mergers achieving efficiency gains or other important performance or wealth-improving gains. Their findings undermine a major rationale for consolidation and in doing so raise questions about other benefits mergers may provide to the public and about alternative motivations such as gaining market power and their effects on the public. Almost all the evidence in the Pilloff and Santomero study is based on data from the 1980s. The 1990s have seen many more mergers of much larger and more geographically widespread banks, and evidence from this decade provides a larger group of mergers to study.

A recent empirical study by Pilloff (1996) covers only the first two years of this decade, but it hints at the motivations for mergers in the 1990s. Pilloff builds upon the work of Cornett and Tehranian (1992), who use data on thirty acquisitions from 1982 to 1987. Those authors examine balance-sheet and stock market data on mergers involving publicly traded banking organizations and find some evidence of superior postmerger performance. Cash flows on assets increased, resulting from the merged institution's enhanced ability to attract loans. Evidence also shows increased employee productivity and asset growth. Cornett and Tehranian examine accounting information for evidence of performance improvements resulting from large banking organization mergers and compare the findings with earnings expectations for such acquisitions using stock market data and standard event study methodology.

Pilloff (1996) studies forty-eight deals, again between publicly traded banking firms or their subsidiaries, over the 1982–91 period. He concludes that, while profitability appears to be unchanged, interpretation of the efficiency effects is less clear. The ratio of postmerger expenses to assets increases but so does return on equity, reflecting the fact that both revenues and leverage increase to offset the increase in expenses. The implication is that operating efficiency decreases slightly, despite the fact that returns to shareholders rose. Finally, there is some evidence of a slight increase in the resultant bank's loan-to-asset ratio.

Cross-sectional analysis of the merging banks' characteristics implies that differences in the premerger characteristics of the acquiring and acquired firms do not affect the resultant firm's postmerger performance, but the premerger characteristics of the acquiring firm do affect postmerger performance. Specifically, acquisitions by less-efficient firms are more likely to result in efficiency gains as measured by expense variables. Acquisitions by larger acquirers and of relatively larger targets are more likely to be associated with reduced expenses but not necessarily with lower total costs. Finally, despite the claims of acquirers, greater overlap of the merging banks' markets does not seem to be associated with better performance. This result brings into question the often-claimed expected benefit that the ability to reduce marketing departments or eliminate redundant offices would result in improved efficiency or profitability.

In terms of the equity market's view of these acquisitions, Pilloff's evidence suggests that the market, on average, does not expect improved profit performance. However, his cross-sectional regressions reveal that improved performance is more likely to be expected in cases involving high premerger expense ratios, especially when the acquired firm has a higher expense ratio than the acquiring firm and the merger partners have a high degree of market overlap.

More recently, Banerjee and Cooperman (1998) use event study methodology to investigate returns to targets and acquirers, using data on a sample of thirty acquiring and sixty-two target banking organizations with assets greater than $100 million between the years 1990 and 1995. The methodology presumes that equity investors are well informed about the average impact of mergers on profit expectations and about the characteristics of the specific merger partners that may affect the results of their particular deal. They compare econometrically the stock market performance of the merging banks' shares with the average control performance during several periods leading up to the merger. These periods are called event windows and begin from one to fifty days before the merger was announced and extend through the actual announcement day. The differences between the merging banks' predicted and actual share performance compared with that of a control sample of nonmerging banks during the event windows are called abnormal returns.

With the breakdown of McFadden Act barriers to interstate mergers, many more potential acquisitions became legally and practically feasible than under the regional compact regime in place during the late 1980s and early 1990s.

Banerjee and Cooperman find a significantly negative abnormal cumulative return of 1.3 percent in a one-day window, [–1, 0], for the acquiring firms, and a huge and significantly positive abnormal return—13.11 percent—for target firms over this same period. They also find significant cumulative abnormal returns for the target for every event window from [–50, 0] to [–1, 0], and an 11.3 percent abnormal return on the day following the merger announcement. Acquiring firms' returns were slightly less negative and marginally statistically significant on the day of the merger and the day following. The authors then specify four hypotheses to explain these abnormal returns and the motives for mergers. These are an efficiency hypothesis, a capital-quality hypothesis, a risk-reduction hypothesis, and a profitability hypothesis. Their efficiency hypothesis suggests, ceteris paribus, that when the target firms are relatively less efficient than the acquirers, the merger will offer the combined firms greater opportunity to realize increased profitability through efficiency gains, thus generating higher abnormal returns. Their capital-quality hypothesis suggests that acquirers with higher capital ratios will experience greater abnormal returns. To explain the relative distribution of abnormal returns between target and acquirer, the authors suggest that

targets with lower capital ratios relative to the acquirer will have higher abnormal returns. The remaining hypotheses suggest that the less efficient, the more profitable, and the less risky the target is relative to the acquirer, the higher the abnormal returns to the target will be. Using cross-section regressions to test these hypotheses, they find some support for the efficiency hypothesis, no support for the capital-quality hypothesis, some support for the profitability hypotheses, and weak support for the risk-reduction hypothesis.

Consolidation in the 1990s

This article presents evidence on efficiency gains and other impacts of large bank mergers, mostly in the 1990s. It replicates the analysis of Pilloff (1996) and investigates market reactions to these mergers. The principal difference between this article and the others cited is that the sample of mergers is both more current and larger. It consists of consolidations among traded firms occurring in the period from 1989 to 1996.[1] While the period includes three years' overlap with the Pilloff study and six years' overlap with that of Banerjee and Cooperman, it does encompass more recent acquisitions as well. Not only is this sample of traded firms larger but it also includes a number of consolidations among the largest banking organizations in the country. Moreover, with passage of the Reigle-Neal Act of 1994 and the resulting breakdown of the McFadden Act barriers to interstate mergers, many more potential acquisitions became legally and practically feasible than under the regional compact regime in place during the late 1980s and early 1990s.

Data. The data in this study include all mergers between publicly traded firms occurring between 1989 and 1996. As in Pilloff's study, the following restrictions are required of the sample: (1) both the acquired and acquiring firms have daily return data available on the Center for Research on Securities Prices (CRSP) files, or their parent companies do; (2) the merger is the primary transaction for the acquired and acquiring companies one year before and one year after the announcement date; (3) for at least three months prior to and thirty days after the acquisition, no other mergers of either firms were announced; (4) neither firm was encouraged by its primary banking regulator to seek a merger partner; and (5) there must have been at least eight quarters of pre- and postmerger performance data available on the surviving acquirer. Information on returns and the market index used to calculate the abnormal returns comes from the daily CRSP files. Company income and balance-sheet data are from the Federal Reserve Y9 Bank Holding Company Report and the Report of Condition and Report of Income and Dividends. Finally, data on deal characteristics and timing are from the SNL Securities database.

Table 2 summarizes the key features of the sample on a year-by-year basis. The mean asset size of acquirers over the period is $32 billion while the mean target is $5 billion. These sizes are substantially larger than the mean acquirer size of $5 billion and mean target size of $3 billion in Pilloff's sample. However, the relative size of the target in the current sample, 16 percent, is smaller than the relative target size in Pilloff's sample, which was 26 percent.[2]

Performance Changes. To investigate the performance effects of mergers, consolidated pre- and post-merger ratios of profitability, operating efficiency, and portfolio composition are compared. For bank holding companies all comparisons are on a consolidated entity basis, and for bank acquirers the comparisons are on a bank basis. As a benchmark and to abstract from industry-wide effects, each performance ratio was calculated on an industry-adjusted basis. Following Pilloff, the change-in-performance ratio, $\Delta X(j)$, is the difference between the pro forma industry-adjusted measure for the institutions involved in the merger, X^{pre}, and the postmerger industry-adjusted measure, X^{post}. For each institution, X^{pre} is calculated as the difference between the average of the eight quarters prior to the merger less the average for a peer group of institutions. Similarly, X^{post} is the average of the eight quarters after the acquisition. To construct the peer or benchmark industry comparison, Pilloff used all firms in excess of $1 billion in assets and formed six geographic subregions for the nation as a whole. Because of the extensive changes in the geographic span of banking organizations during the 1990s as restrictions on interstate banking have been phased out, geographic controls are now less meaningful than they were for Pilloff's sample. This article does not report tests of merging banks' performance variables on the basis of geographic categories.

Table 3 compares the pro forma pre- and post-merger performance and change in performance for the sample. The pro forma merging institutions appear to have higher earnings-to-asset ratios but a lower rate of return on equity and higher noninterest expense than the control firms. They are better capitalized and somewhat more efficient, having a lower ratio of expenses to assets and a lower ratio of expenses to revenues; and they make more loans than their peers. After the merger, the resulting institutions remain slightly better earners in terms of their rate of return on assets but the income-to-equity ratio declines somewhat more. They also continue to have statistically significantly lower total expenses, higher noninterest ex-pense, and less leverage; and they make more loans than their peer institutions.[3]

However, looking at the change in the performance measures, it is not clear that mergers result in signifi-

cant performance improvements. Overall performance may even deteriorate. Two of the three measures of rate of return on equity decline significantly. Expense efficiency is mixed, with the expenses-to-asset ratio increasing somewhat relative to peer institutions, while the expense-to-revenue ratio remains below that of peer institutions. Finally, the ratio of premises to assets is larger, and there is a marginally statistically significant decrease in leverage while the loan-to-asset ratio increases significantly.

These results are different from those found in Pilloff's earlier study. Pilloff finds an efficiency improvement in terms of the ratio of expenses-to-revenue and observes that postmerger leverage increases relative to the peer institutions whereas leverage decreases in the 1990s data. He observes only two of the six profitability measures—both measures of the rate of return on equity—to be significantly different after merger, and this effect is the result of increased leverage, which increases profitability.[4] In addition, the combined pro forma banks are less different from their peers in Pilloff's study than in this more recent sample. Pilloff also finds smaller loan-to-asset ratio effects, albeit in the same direction, but the change in the ratio is not statistically significant.

Postmerger performance may be influenced by the premerger performance of either the acquirer or target or the relative difference in acquirer and target firm performance. For example, a strong acquirer may believe that it has superior managerial capabilities and thus look for poor performing targets to which its superior management may be applied. Alternatively, a poorly performing acquirer may seek a merger partner and use the acquisition as the catalyst to overcome managerial inertia and improve its operations.

To test the influence of a merger partner's characteristics, correlations between premerger characteristics and changes in postmerger performance are examined in Table 4. The greater the acquirer's profitability, the more negative the merger's earnings impact is, and the same is true for the target. There is some evidence of a greater efficiency gain the larger the target's expense ratio is. However, the impact is somewhat offset when the acquirer's profits are large relative to the target. The results are not supportive of positive earnings impacts of mergers in general.

Pilloff also hypothesizes that performance changes may be related to both size and the relative size of the acquirer and target. Table 5 presents the correlations between changes in performance measures and the target's and acquirer's initial size and their relative size. Two of the changes in profitability are positively related to acquirer size while only one of the target's profit characteristics is marginally positively significant. This finding may suggest that the larger the acquirer is relative to the target, the more the change in profitability is likely to be positive. Expenses are more likely to increase when the acquirer is large, as is the loan-to-asset ratio. Only the target's initial size is positively related to the change in the resultant bank's capital position. Finally, the larger the relative size of the target is, the more likely leverage is to be reduced and the more likely the loan ratio is to decline.

Overall, the results do not suggest that the 1990s mergers have resulted in either a positive earnings performance or greater efficiency. Only one of the acquirer's profitability ratios is positively related to its size. Otherwise, larger acquirers are more likely to experience increases in the expense-to-assets ratio and an increase in loans. Larger targets are more likely to be associated with a reduction in leverage.

Market Responses to Megamergers. To see how well and whether the market anticipates and prices any of the acquisitions based on the initial characteristics of the targets or acquirers or whether the changes in performance are anticipated and priced, a standard event study (see Dodd and Warner 1983) is performed similar to that of Pilloff. Several event windows are used to calculate abnormal returns ranging in size from twenty-one days, spanning days $[t = -20, t = 0]$ to only two days $[t = -1, 0]$.

In estimating excess returns, a single-factor ordinary least squares market model is used to provide firm-specific adjustments for risk. The CRSP equally weighted market index is used as the proxy for the market. The parameters of the model are estimated over the period from $[t - 300]$ through $[t - 30]$, where the event day of

1. Use of eight quarters of past merger data to evaluate postmerger performance effectively limits the study to mergers consummated no later than year-end 1996.

2. In more than 55 percent of the acquisitions, the acquired firm was 20 percent the size of the acquiring firm or greater. In 20 percent of the cases, the acquired firm was less than 3 percent the size of the acquiring firm. This distribution suggests that many of the acquisitions had the potential to meaningfully impact the performance of the resulting firm.

3. Since the time for X^{post} *starts right after the merger, the costs of the merger are also considered. While this may tend to hide some of the longer-term cost or performance differences, estimates of those costs by equity market participants would be expected to affect the resulting firm's performance from the stockholders' perspective as they value the acquiring firm in the acquisition.*

4. Only one of the changes in profitability measures (ROE3) was statistically significant whereas two of the change measures in Table 3—ROE1 and ROE2—are marginally statistically significant at the 10 percent level.

the merger or acquisition is defined as day [$t = 0$]. The firm-specific, single-index model control return is of the form $\hat{R}_{j,t} = \beta_0 + \beta_1 R_{m,t}$, where $\hat{R}$ is the control for bank j during day t, $R_{m,t}$ is the return on the CRSP equally weighted index, and β_t is estimated model parameters.

Daily abnormal returns for a given bank j on day t are defined as $AR_{j,t} = R_{j,t} - \hat{R}_{j,t}$, where $AR_{j,t}$ is the abnormal return for bank j on day t, and $R_{j,t}$ is the realized daily return for bank j on day t. Individual abnormal returns $AR_{j,t}$ are then aggregated to form a portfolio of daily abnormal returns:

$$AR_t = \frac{1}{N}\sum_{j=1}^{N} AR_{j,t}, \qquad (1)$$

where N is the number of firms in the portfolio for day t. Cumulative abnormal returns are calculated as

$$CAR_t = \sum_{t_b}^{t_c} AR_t, \qquad (2)$$

where t_b begins and t_c ends the cumulating period.

The results are shown in Table 6 and indicate that only the abnormal returns in the window [–1, 0] are statistically significant and positive, amounting to a two-day cumulative abnormal return of 0.77 percent. In contrast, Pilloff finds only a statistically significant cumulative return for the much longer event window [–10, 0], and that return is also larger, at 1.4 percent. Overall, the more current sample has a narrower range of plus and minus abnormal returns for event period [–1, 0] than Pilloff's sample, but it does have a larger standard deviation.[5]

The question then arises whether the abnormal returns are related to the premerger characteristics or sizes of the merging firms or to the changes in their performance ratios. Table 7 shows the correlations between the initial merger partners' ratios and sizes and the abnormal returns for the [–1, 0] event window. Only one each of the acquirer's and target's profitability measures are positive and significant. Moreover, the higher the expense-to-assets ratio, the lower the abnormal returns. The higher the core deposits ratio of the acquirer, the more likely abnormal returns are to be positive. Finally, only the target's size and the relative size of the acquirer to the target are positively related to abnormal returns. Interestingly, in Table 8 none of the abnormal returns were related to performance changes.

Conclusion

Using data from the 1990s to extend existing analysis of banking mergers, this article examines the performance and value effects of banking organization acquisitions. Specifically, examining recent data allows considering whether there is evidence of efficiency or other gains from the wave of in-market and market extension acquisitions flowing from the erosion and final elimination of the McFadden Act.

Consistent with the results of earlier studies, the efficiency and performance effects were mixed. Evidence suggests that the better-performing institutions tended to target the higher-performing targets, but the resulting mergers did not significantly improve profit performance or efficiency. There were marginal declines in leverage and increases in loan portfolio composition. Moreover, the effects, except for portfolio allocation, were even smaller than those found by Pilloff in his study using data primarily from the 1980s. In addition, looking at the market's reaction to proposed mergers, there is only weak evidence that the market viewed acquisitions with favor. It did, however, tend to be less optimistic about the savings from mergers when expense ratios were higher. The overall conclusion is that the widely touted earnings, efficiency, and other performance and earning benefits of megamergers still remain in doubt.

5. Direct comparison with Pilloff's results are difficult because he uses a different method of computing abnormal returns.

TABLE 1 Banking Organization Mergers and Acquisitions, 1989–99

Years	Number of Banks	Bank Assets ($ Millions)	Bank Deposits ($ Millions)
1989	142	81,228	64,035
1990	170	36,392	31,031
1991	262	275,137	202,618
1992	319	104,800	90,315
1993	373	127,529	103,089
1994	444	99,981	79,740
1995	358	486,275	347,684
1996	364	188,727	155,399
1997	346	256,857	196,249
1998	406	1,086,872	656,882
1999[a]	81[a]	87,132	59,605
Total	3,844	3,210,785	2,279,563

[a] As of April 5, 1999

Source: Information on returns and the market index used to calculate the abnormal returns came from the daily CRSP files. Company income and balance-sheet data are from the Federal Reserve Y9 Bank Holding Company Report and the Report of Condition and Report of Income and Dividends. Data on deal characteristics, acquisition prices, and premium calculations are from the SNL Securities database.

TABLE 2 Summary of Mergers in Sample

	Number of Mergers	Total Assets ($ Millions) at End of Quarter before Merger Date						Relative Size (Percent)[a]
		Acquirer			Target			
		Mean	Minimum	Maximum	Mean	Minimum	Maximum	Mean
1989	4	18,867	5,730	48,048	1,016	273	2,308	9.2
1990	8	20,511	470	48,737	3,010	261	7,883	19.3
1991	7	22,478	384	76,884	10,672	314	65,639	24.7
1992	18	28,380	1,289	119,902	6,751	207	72,873	16.7
1993	16	32,254	511	141,966	2,509	95	16,626	10.0
1994	18	36,064	519	197,543	2,262	154	21,643	12.9
1995	9	23,671	283	83,803	5,855	204	33,122	22.3
1996	14	53,020	1,200	194,375	9,319	177	54,593	18.0
Full Sample	94	32,216	283	197,543	5,196	95	72,873	16.1

[a] Relative size equals target total assets divided by target plus acquirer total assets, with assets measured at the end of the quarter before the merger date.

Sources, Tables 2-8: Information on returns and the market index used to calculate the abnormal returns came from the daily CRSP files. Company income and balance-sheet data are from the Federal Reserve Y9 Bank Holding Company Report and the Report of Condition and Report of Income and Dividends. Data on deal characteristics, acquisition prices, and premium calculations are from the SNL Securities database.

TABLE 3 Premerger, Postmerger, and Changes in Performance (Percent)

Performance Measure (X)	Mean X^{pre}	Mean X^{post}	Mean ΔX	Tenth Percentile of ΔX	Ninetieth Percentile of ΔX	Standard Deviation of ΔX
ROA1 = Net Income/Total Assets	0.07***	0.04	–0.03	–0.20	0.18	0.26
ROA2 = Net Operating Income Plus Provisions/Total Assets	0.11	0.08**	–0.04	–0.30	0.27	0.35
ROA3 = Net Operating Income Less Provisions/Total Assets	0.05*	0.05*	0.00	–0.20	0.27	0.23
ROE1= Net Income/Total Equity	0.22	–0.52	–0.74*	–3.70	1.92	0.04
ROE2 = Net Operating Income Plus Provisions/Total Equity	0.53	–0.52	–1.05*	–4.60	3.03	5.46
ROE3 = Net Operating Income Less Provisions/Total Equity	–0.95**	–1.09**	–0.14	–4.50	3.15	3.25
EXPAST = Expenses/Total Assets	–0.28***	–0.19***	0.09**	–0.30	0.65	0.37
EXPREV = Expenses/Revenue	–1.55***	–1.89***	–0.34	–4.80	4.22	3.46
SALAST = Salaries/Total Assets	0.04*	0.05**	0.01	–0.10	0.14	0.11
PREMAST = Premises/Total Assets	0.01	0.02*	0.01**	–0.00	0.07	0.05
NNIXAST = Noninterest Expense/Total Assets	0.10**	0.12**	0.01	–0.20	0.33	0.24
EQAST = Equity/Total Assets	0.68***	0.87***	0.19*	–0.70	1.20	1.07
LOANAST = Loan/Total Assets	6.11***	7.82***	1.72**	–6.00	10.96	7.10
CORAST = Core Deposits/Total Assets	9.95***	9.50***	–0.45	–6.90	5.34	6.26

Note: The term X^{pre} is the average performance during the eight quarters preceding the merger for the target and acquirer combined, and X^{post} is the average combined performance during the eight quarters following the merger. The term ΔX is the difference between premerger and postmerger performance. All performance measures control for size. The notations *, **, *** indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

TABLE 4
Correlations of Performance Changes with Premerger Performance Variables

Performance Measure (X)	Corr (ΔX, X^R)	Corr (ΔX, X^{RA})	Corr (ΔX, X^{RT})
ROA1 = Net Income/Total Assets	0.342***	–0.226**	–0.480***
ROA1 = Net Operating Income Plus Provisions/Total Assets	0.315***	–0.234**	–0.479***
ROA1 = Net Operating Income Less Provisions/Total Assets	0.287***	–0.181*	–0.450***
ROE1 = Net Income/Total Equity	0.363***	–0.252**	–0.463***
ROE2 = Net Operating Income Plus Provisions/Total Equity	0.344***	–0.307***	–0.494***
ROE3 = Net Operating Income Less Provisions/Total Equity	0.057	–0.402***	–0.385***
EXPAST = Expenses/Total Assets	0.255**	0.135	–0.223**
EXPREV = Expenses/Revenue	0.342***	–0.149	–0.440***
SALAST = Salaries/Total Assets	0.315***	0.284***	–0.096
PREMAST = Premises/Total Assets	0.168	0.169	–0.028
NNIXAST = Noninterest Expense/ Total Assets	0.201*	0.152	–0.098
EQAST = Equity/Total Assets	0.239**	0.146	–0.113
LOANAST = Loans/Total Assets	0.234**	0.085	–0.180*
CORAST = Core Deposits/Total Assets	0.261**	–0.088	–0.254**

Note: The term ΔX is the difference between premerger and postmerger performance. The term X^R is the weighted difference between acquirer and target premerger performance. The term X^{RA} is the weighted measure of acquirer premerger performance, and X^{RT} is the weighted measure of target premerger performance. All performance measures control for size. The notations *, **, *** indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

TABLE 5 Correlations of Performance Changes with Size

Performance Change (ΔX)	Corr (ΔX, LNAAST)	Corr (ΔX, LNTAST)	Corr (ΔX, RELSIZE)
ΔROA1	0.156	0.117	–0.060
ΔROA2	0.218**	0.180*	–0.048
ΔROA3	0.097	0.143	0.053
ΔROE1	0.139	0.057	–0.124
ΔROE2	0.186*	0.096	–0.124
ΔROE3	0.054	–0.011	–0.105
ΔEXPAST	0.267***	0.055	–0.188*
ΔEXPREV	–0.087	–0.158	–0.078
ΔSALAST	–0.116	–0.147	–0.034
ΔPREMAST	0.148	0.072	–0.088
ΔNNIXAST	0.009	–0.044	–0.074
ΔEQAST	0.063	0.212**	0.243**
ΔLOANAST	0.340***	0.127	–0.250**
ΔCORAST	–0.074	–0.056	–0.022

Note: The term ΔX is the difference between premerger and postmerger performance. The terms LNAAST and LNTAST are the logs of acquirer's and target's total assets. Relative size equals target total assets divided by target plus acquirer total assets. Total assets are measured at the end of the quarter before the merger date. All performance measures control for size. The notations *, **, *** indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

TABLE 6 Consolidated Abnormal Returns at Merger Announcement (Percent)

Event Window	Mean	Tenth Percentile	Ninetieth Percentile	Standard Deviation
[–1, 0]	0.77*	–3.00	4.93	4.18
[–2, 0]	0.57	–3.60	4.73	4.08
[–5, 0]	0.56	–3.90	4.36	5.09
[–7, 0]	0.35	–4.60	4.50	5.25
[–10, 0]	–0.01	–4.90	4.52	5.04
[–15, 0]	0.01	–5.90	5.09	5.53
[–20, 0]	0.28	–6.20	6.23	5.89

Note: Consolidated abnormal returns equal the cumulative weighted realized returns of acquirers and targets less the cumulative weighted expected returns of acquirers and targets during the event window, with the announcement date at day 0. Expected returns are calculated from a standard market model. The notation * indicates significance at the 10 percent level.

TABLE 7
Correlations of Consolidated Abnormal Returns at Merger Announcement with Premerger Variables

Performance Measure (X)	Corr (CAR, X^R)	Corr (CAR, X^{RA})	Corr (CAR, X^{RT})
ROA1 = Net Income/Total Assets	–0.070	0.140	0.137
ROA2 = Net Operating Income Plus Provisions/Total Assets	–0.094	0.082	0.140
ROA3 = Net Operating Income Less Provisions/Total Assets	–0.212**	0.046	0.227**
ROE1 = Net Income/Total Equity	–0.038	0.219**	0.122
ROE2 = Net Operating Income Plus Provisions/Total Equity	–0.062	0.139	0.130
ROE3 = Net Operating Income Less Provisions/Total Equity	–0.211**	0.110	0.271**
EXPAST = Expenses/Total Assets	–0.139	–0.554***	–0.300***
EXPREV = Expenses/Revenue	0.109	–0.144	–0.192*
SALAST = Salaries/Total Assets	0.026	–0.023	–0.057
PREMAST = Premises/Total Assets	0.031	0.040	0.003
NNIXAST = Noninterest Expense/ Total Assets	0.038	–0.011	–0.057
EQAST = Equity/Total Assets	–0.095	–0.050	0.056
LOANAST = Loans/Total Assets	0.044	0.188*	0.125
CORAST = Core Deposits/Total Assets	0.246**	0.257**	0.034

Premerger Variable (Z)	Corr (CAR, Z)
LNAAST	0.116
LNTAST	0.232**
RELSIZE	0.194*

Note: The term CAR equals the cumulative weighted returns of acquirers and targets less the cumulative weighted expected returns of acquirers and targets from one day before to the day of the merger announcement. Expected returns are calculated from a standard market model. The term X^R is the weighted difference between acquirer and target premerger performance. The term X^{RA} is the weighted measure of acquirer premerger performance, and X^{RT} is the weighted measure of target premerger performance. All performance measures control for size. The notations *, **, *** indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

TABLE 8
Correlations of Consolidated Abnormal Returns at Merger Announcement with Performance Changes

Performance Change (ΔX)	Corr (CAR, ΔX)
ΔROA1	–0.106
ΔROA2	–0.047
ΔROA3	–0.007
ΔROE1	–0.149
ΔROE2	–0.104
ΔROE3	–0.147
ΔEXPAST	–0.171
ΔEXPREV	–0.126
ΔSALAST	–0.032
ΔPREMAST	0.001
ΔNNIXAST	–0.012
ΔEQAST	0.131
ΔLOANAST	0.082
ΔCORAST	0.117

Note: The term CAR equals the cumulative weighted returns of acquirers and targets less the cumulative weighted expected returns of acquirers and targets from one day before to the day of the merger announcement. Expected returns are calculated from a standard market model. The term ΔX is the difference between premerger and postmerger performance. All performance measures control for size.

REFERENCES

BANERJEE, AJEYO, AND ELIZABETH COOPERMAN. 1998. "Returns to Targets and Acquirers: Evidence for Bank Mergers in the '90s." Unpublished working paper, University of Colorado, Denver.

CORNETT, MARCIA MILLON, AND SANKAR DE. 1991. "Common Stock Returns in Corporate Takeover Bids: The Evidence from Interstate Banking." *Journal of Banking and Finance* 15 (April): 273–95.

CORNETT, MARCIA MILLON, AND HASSAN TEHRANIAN. 1992. "Changes in Corporate Performance Associated with Bank Acquisitions." *Journal of Financial Economics* 31 (April): 211–34.

DODD, PETER, AND JEROLD B. WARNER. 1983. "On Corporate Governance." *Journal of Financial Economics* 11:401–38.

PILLOFF, STEVEN J. 1996. "Performance Changes and Shareholder Wealth Creation Associated with Mergers of Publicly Traded Banking Institutions." *Journal of Money, Credit, and Banking* 28 (August): 294–310.

PILLOFF, STEVEN J., AND ANTHONY M. SANTOMERO. 1997. "The Value Effects of Bank Mergers and Acquisitions." Financial Institutions Center, The Wharton School, University of Pennsylvania, Working Paper No. 97-07.

Economic Policy Trends in Post–World War II Latin America

CARLOS LOZADA
The author is an associate editor of Foreign Policy *magazine. This article was completed while he was an economic analyst in the macropolicy section of the Atlanta Fed's research department.* *He thanks Nancy Birdsall for collaboration on an earlier version of this article and Frank King, Michael Chriszt, and Mary Rosenbaum for helpful comments.*

RECENT ECONOMIC DISTURBANCES IN LATIN AMERICA—PARTICULARLY THE CURRENCY CRISES IN MEXICO IN 1994–95 AND BRAZIL IN 1998–99—HAVE PROMPTED SIGNIFICANT RESEARCH AND DEBATE OVER FINANCIAL SECTOR REFORMS AND APPROPRIATE MONETARY AND FISCAL POLICY FOR THE REGION. THE RECENT DISCUSSION OVER DOLLARIZATION IS BUT ONE OF MANY SUCH DEBATES. IT IS IMPORTANT, HOWEVER, THAT THE ONGOING POLICY DISCOURSE BE INFORMED BY A BROADER UNDERSTANDING OF THE REGION'S ECONOMIC HISTORY.[1]

The current rethinking of economic policy in Latin America is only the latest chapter of a much longer story. Well before the recent episodes of financial turmoil, Latin American economies had already proven vulnerable to external economic shocks, which have taken the form of changes in commodity prices, movements in international interest rates, and fluctuations in the volume and direction of capital flows. These factors have interacted with (and in some cases, prompted) frequent changes in the region's economic policy orientation. This mix has resulted in high volatility of key indicators, including inflation, fiscal and external balances, and gross domestic product (GDP) growth rates. This article provides a survey of the evolution of economic policy and performance in Latin America in the post–World War II period. It highlights the impact of and reaction to certain economic shocks the region experienced, including the declining terms of trade in the early postwar period, the oil shocks of the 1970s, the debt crisis of the 1980s, and the more recent emerging markets crises of 1997–99.

Prebisch and Import Substitution

The notion that export-led development was the wrong choice for Latin America took hold in the immediate aftermath of World War II. The international recession of the 1930s, the economic turmoil caused by the global conflicts, and protectionist policies by developed nations such as the U.S. Smoot-Hawley tariff in 1930 led to weak demand for primary commodities and consequent contractions for Latin American economies, which have traditionally been heavily dependent on primary exports. The perception of a global division of labor, with the north producing manufactured items and the south providing primary goods, seemed inimical to Latin America's long-term develop-

ment because of adverse terms of trade fluctuations and to the apparent concentration of technology in the manufacturing industries of the north. This north-south or "center-periphery" dichotomy, forcefully articulated by Raul Prebisch of the United Nations' Economic Commission for Latin America and the Caribbean (CEPAL), came to dominate regional economic thought during the early postwar years.

Economic policy making in Latin America was guided by the principle that the international environment presented an obstacle to the region's economic expansion and that policy should be adjusted to deal with external constraints on growth. Prebisch (1963) argued that domestic industrialization would foster the spread of technology, increase employment, and enhance the productivity of the labor force, thus reducing the region's vulnerability to international economic forces. This argument served as the underlying rationale for the region's import-substitution industrialization (ISI) policies, which sought to enhance industrial development through the protection of domestic markets via tariffs, quotas, and other restrictions and with targeted subsidies to local producers.

In its early period, import substitution was able to foster heavy industries in some of the larger countries of the region and created a modest base for the growth of domestic manufacturing. As a result, Latin American GDP grew at an average annual rate of 5.1 percent from 1951 to 1960 (Wilkie 1995). However, the strategy began encountering bottlenecks in the late 1960s. First, the production of advanced durable goods often required intermediate capital inputs not available domestically. This need for foreign inputs aggravated the very problem that the region was attempting to avoid: external dependence. Also, in order to exploit economies of scale, complex goods often required larger markets than those available internally.

Import substitution also resulted in the creation of capital-intensive industries, thus failing to generate a substantial demand for labor. Although some high-productivity manufacturing jobs were created, the expected connection between manufacturing, technology, and increased labor force productivity with subsequent higher wages and living standards failed to materialize. But despite the strategy's shortcomings, the regional economy experienced even higher growth rates in the 1960s, expanding by an average annual rate of 5.75 percent from 1961 to 1973 (Wilkie 1995). The results appeared favorable enough that the trend of protectionism and government intervention in Latin America continued during the 1970s.

To continue the growth process, many Latin American countries began importing heavily, relying on capital inflows greatly facilitated by the "petrodollars" from the oil shock of 1973 to complement internal savings in the financing of investment. The significant liquidity available in the international system made cheap foreign financing readily available during the 1970s. Between 1975 and 1982, Latin America's long-term foreign debt increased from $43 billion to $176 billion (Edwards 1993). In keeping with the top-heavy nature of ISI, state enterprises and the state in general received the lion's share of funds as the inflows financed ever-increasing public sector deficits (Kuczynski 1988). A favorable international environment of low real interest rates and strong demand for Latin American primary exports helped the region's governments service their growing debts with little difficulty from 1975 through 1979.

During most of the economic crises of the post–World War II period, governments in Latin America responded by introducing new economic policies representing a significant departure from the prior policy path.

The international environment eventually became less accommodating. The second oil shock in 1979 resulted in higher petroleum prices and declines in the prices of other primary commodities. Meanwhile, to combat inflationary pressures, U.S. monetary authorities raised interest rates, significantly increasing Latin America's debt-service burden. Because much of the debt Latin American governments had assumed was in the form of variable-rate loans, interest payments on the region's foreign debt rose significantly, from less than $9 billion in 1978 (17 percent of regional export earnings) to $30 billion in 1981 (42 percent). The region's current account deficits more than doubled in the 1979–81 period (Inter-American Development Bank 1985). These factors, in conjunction with a pattern of capital flight from Latin America during the late 1970s, eventually triggered the regional debt crisis of the 1980s.

The Lost Decade and the Return to the Market

In 1982 Mexican authorities declared themselves unable to continue servicing the country's external obligations; other economies in the region quickly followed suit. Borrowing from abroad plummeted from $48 billion in 1981 to $16 billion in 1983, and capital

1. This article builds upon Birdsall and Lozada (1998).

inflows virtually disappeared by mid-decade (Inter-American Development Bank 1985).

Faced with a sudden and drastic decline in capital inflows, many Latin American countries restricted imports, imposed higher tariffs, created multiple exchange rates, and levied inefficient taxes in hope of mitigating ballooning external and domestic deficits. As deficits persisted and foreign funding continued declining, governments turned to their central banks for financing, provoking inflation. In the mid-1980s, some countries experimented with heterodox strategies, attempting to combat "inertial" inflation through price freezes as well as exchange rate and wage controls. These approaches ultimately exacerbated the inflationary spiral, mainly because they neglected to recognize the fundamental role of fiscal discipline in achieving price stability. Inflation reached nearly 7,500 percent in Peru during 1990, and several other countries — notably Argentina, Bolivia, Brazil, and Nicaragua—experienced bouts of hyperinflation at some point during the decade. These policy changes contributed to poor economic growth rates during the decade. After averaging 5.9 percent during the 1970s, annual GDP growth rates dropped to 1 percent from 1980 to 1990 and were negative on a per capita basis during that period (Inter-American Development Bank 1996).

Unlike the earlier crisis episodes, which prompted a sea change in the economic policy regime of several economies, the recent external shocks have led most governments to deepen, not depart from, the overall market-oriented framework.

By the late 1980s a new consensus on economic policy was slowly emerging. Policymakers began to recognize that the state-led, protectionist development model employed over the previous decades had finally exhausted itself. The visible success of Chile, an early adjuster, the collapse of the statist economies of Eastern Europe and the Soviet Union, and the then-rapid growth of East Asian economies all encouraged Latin American policymakers to adopt market-based reforms: greater openness to international trade, strict fiscal discipline, and privatization of state-owned enterprises. The policy shift also reflected the growing influence of U.S.-trained Latin American economists at the technical and political level in several countries and the influence of the multilateral financial institutions in the form of loans, empirical studies, and policy dialogue. The magnitude of the crisis in the 1980s may also have contributed to the policy shift as the region's deteriorating economic conditions expanded the political space available to Latin American leaders and made radical policy changes viable.

While the scope and pace of policy reform varied significantly across countries (Rosenthal 1996), some regional generalizations are possible. In the countries that carried out comprehensive adjustments, the initial reforms were aimed at stabilizing the economy, that is, curtailing inflation. These included greater fiscal discipline and tight monetary policy and, in some cases, the use of nominal exchange rate anchors. The role of the state, so prominent in previous decades, was diminished in favor of the private sector and market forces. Between 1988 and 1993, Latin America accounted for more than half the total value of worldwide divestitures, surpassing both Asia and Eastern Europe (Birdsall, Graham, and Sabot 1998). Deregulation and privatization of state enterprises fostered greater competition and helped reduce distortions in the economy. Finally, many countries unilaterally cut tariffs and eliminated various barriers to trade, helping accelerate regional productivity growth. Average tariff rates in Latin America declined from more than 50 percent in 1985 to approximately 10 percent a decade later (Birdsall, Graham, and Sabot 1998). The creation of trade arrangements such as the North American Free Trade Agreement (NAFTA) and the Common Market of the South (MERCOSUR) also contributed to increases in regional trade.

These policies stood in sharp contrast to the consensus of the early postwar period, under which external factors had been considered the most significant impediments to growth. Now, rather than being the solution to the external constraints, as Prebisch had advocated,protective and state-centered domestic policies were seen as the root cause of the region's economic underperformance. The view that impediments to growth could be removed by eliminating internal policy distortions came to dominate economic thought within the region and among multilateral institutions. The Inter-American Development Bank and the World Bank both emphasized the primacy of domestic policies early in the 1990s: "The progress of the Latin American economies during the coming years is likely to be driven more by the success of domestic reform processes than by the performance of the world economy" (Inter-American Development Bank 1992); "[t]he future of the developing countries is largely in their own hands. . . . The right strategy for the developing countries, whether external conditions are supportive or not, is to invest in people, including education, health, and population control; help domestic markets to work well by fostering competition and investing in infrastructure; liberalize trade and foreign investment; avoid excessive fiscal deficits and high inflation" (World Bank 1991).

TABLE 1 Annual CPI Inflation in Selected Latin American Economies (1991–99)

	1991	1992	1993	1994	1995	1996	1997	1998	1999f
Argentina	171.7	24.9	10.6	4.2	1.6	0.0	0.3	0.7	-1.0
Brazil	440.9	1,008.7	2,148.5	2,668.6	23.2	10.0	4.8	-1.8	6.6
Chile	22.0	15.4	12.7	11.4	8.2	6.6	6.0	4.7	4.0
Colombia	30.4	27.0	22.6	23.8	19.3	21.6	17.7	16.7	12.4
Ecuador	48.7	54.6	45.0	27.3	22.8	25.5	30.5	43.4	55.5
Mexico	22.7	15.5	9.7	6.9	52.0	27.7	15.7	18.6	14.4
Peru	409.5	73.6	48.6	23.7	10.2	11.8	6.5	6.0	5.7
Venezuela	34.2	31.4	38.1	60.8	56.6	103.2	37.6	29.9	25.3

f = forecast

Source: Inter-American Development Bank data and Latin American Consensus Forecasts

Positive medium-term results vindicated Latin America's market reforms. Stabilization efforts resulted in a significant decline of regional inflation (Table 1). By 1997 most countries in the region had only single-digit inflation, and the median rate had dropped to 9 percent, the lowest in any year since 1977 (Inter-American Development Bank 1998). Similarly, regional economic growth recovered during the first half of the 1990s and reached 5.1 percent in 1994. Although in some countries, such as Argentina, El Salvador, and Peru, strong growth was partially enabled by the renewed employment of previously underutilized productive capacity, in most countries the rapid growth of the early 1990s was fueled by new investments and productivity gains (Inter-American Development Bank 1996).

Capital Flows and Financial Contagion

Much of the new investment in the region was facilitated by foreign capital inflows, which made a strong comeback to the region during the 1990s. After virtually disappearing from 1983 to 1990, financial flows to Latin America grew dramatically, reaching 4 percent of regional GDP in 1991 and 6 percent in 1993 and 1994. Chart 1 shows the evolution of capital inflows to the region from 1988 through 1997.

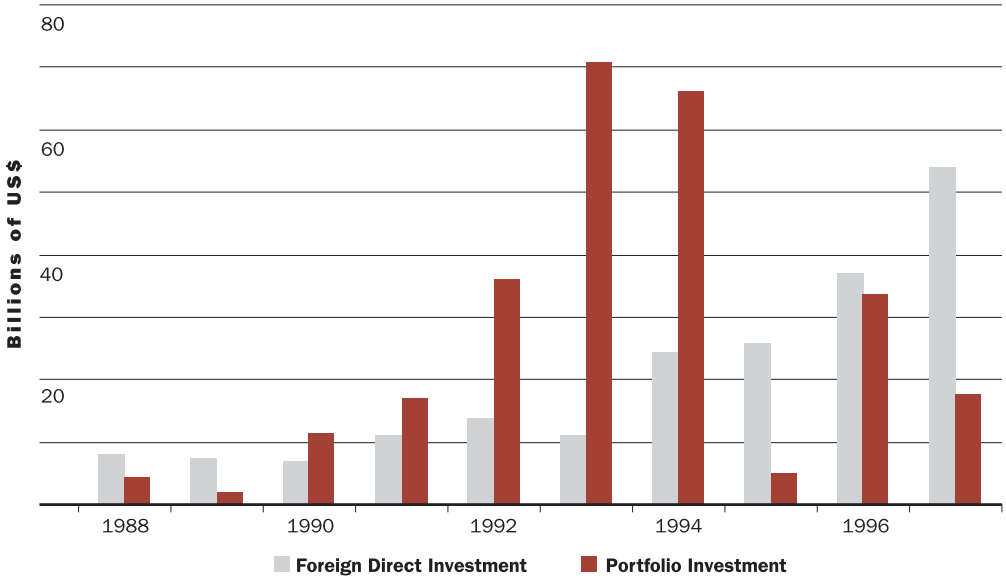
Several factors contributed to this surge of capital inflows. Reductions in inflation, as well as the privatization and deregulation of internal markets, provided domestic and international investors with new high-return opportunities in a less volatile economic environment. The lure of high returns and opportunities for risk diversification supported this increased participation of emerging economies in global capital markets. Finally, an environment of low interest rates in the United States during the 1990s likely rendered the high returns in emerging markets even more attractive.

Furthermore, technological advances greatly increased the pace and efficiency with which capital can flow into and out of emerging markets. High-tech innovations in the global financial system have facilitated the virtually instantaneous dissemination of information among market actors, thus dramatically reducing transaction costs and enabling the development of new financial instruments. The integration of emerging economies into this system allows market forces to swiftly reward positive economic performance or policies as well as to expose and punish underlying economic mismanagement or inconsistencies.

Capital inflows in the context of increasingly integrated financial systems present distinct benefits to emerging economies. Primarily, they can boost economic growth by decoupling investments in the local economy from the availability of domestic savings. Financial integration also supports growth by shifting the investment mix in a given country toward higher-return projects because it improves investors' ability to diversify away some of the risk normally involved in high-return investments (World Bank 1997).

However, increased capital flows and financial integration also present substantial risks (Chang 1999). High levels of capital inflows, common in the early stages of financial market integration, increase the potential costs of sudden reversals. Rapid outflows are usually triggered by a loss of confidence in a country's ability to earn enough to pay its foreign-denominated debt although they might also occur, at least temporarily, in other economies that do not necessarily share the vulnerabilities of the originating nation. Problems in a particular emerging economy might trigger a temporary withdrawal from several or all economies considered to be in the same asset class. Some argue that such outflows may be compounded by a so-called herding effect, by which institutional investors and fund managers tend to follow the behavior of their peers so that their particular portfolio record will not appear worse than that of the industry as a whole.

CHART 1 Investment Flows into Latin America (1988–97)



Source: Inter-American Development Bank

Underperformance is often penalized more than overperformance is rewarded, especially if emerging market fund managers are required to meet the performance of the median fund in their category or market (World Bank 1997).

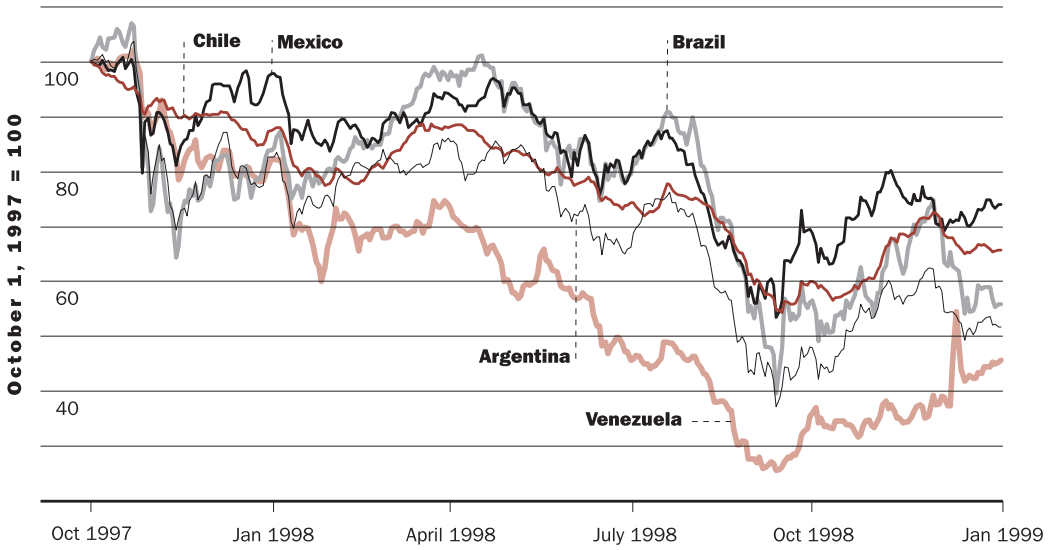
The risks associated with increased capital flows to an emerging economy can be exacerbated when the inflows are intermediated through an underdeveloped or inadequately regulated local financial system, which might already be experiencing a lending boom such as often happens in the early stages of a poststabilization economic recovery. Unless lending practices are rigorously supervised, it is likely that funds are made available to individuals or entities that may be unable to meet their payments in face of a sudden economic downturn or interest rate hike (Birdsall, Gavin, and Hausmann 1998). Unfortunately, such financial fragilities may become apparent only after the economy or the capital inflows or both begin to decelerate.

The Mexican Crisis. This last scenario describes some elements of the Mexican financial crisis of 1994–95. Along with the country's stabilization efforts begun in the late 1980s, authorities had introduced financial sector reforms that included the liberalization of interest rates and the privatization and deregulation of the banking system. These conditions combined to unleash a bank lending boom in the early 1990s. At the same time that capital flows to Mexico grew to an annual average of 8 percent of GDP from 1990 to 1993, bank credit to the private sector rose from less than 10 percent of GDP in 1989 to about 40 percent by 1994 (Birdsall, Gavin, and Hausmann 1998).

Domestic disturbances—both economic and political—prompted a severe devaluation of the Mexican peso in late December 1994; on December 20–21 alone the central bank spent approximately $4 billion in foreign reserves attempting to defend the currency (International Monetary Fund 1995). Authorities were eventually compelled to move to a market-determined exchange rate regime. The crisis was heightened by the fact that substantial stocks of dollar-denominated, short-term government debt was coming due just as international capital markets were least willing to finance Mexico. The impact on the banking sector was significant as inflation, high interest rates, reduced economic activity, increased debt burdens, and deteriorating loan portfolios contributed to the deterioration of banks' capital ratios. The government was eventually compelled to intervene through a series of bailout packages and banking system reform initiatives. Although economic growth in Mexico resumed strongly in 1996 and 1997, the crisis exacted a high price in 1995 when unemployment and bankruptcies rose dramatically and Mexican GDP contracted by 6.2 percent (Inter-American Development Bank 1997).

The Mexican crisis also had regional contagion effects, triggering capital reversals in other Latin American economies in the region, most notably Argentina, which came under pressure because of its "convertibility program," a fixed exchange rate regime that pegs the value of the local peso to the U.S. dollar and prohibits the issuance of any new currency not backed by international reserves. Over a period of three months in early 1995, approximately 18 percent of

CHART 2 Latin American Stock Markets (October 1997–December 1998)



Source: Compiled by the Federal Reserve Bank of Atlanta using data from *Emerging Markets Companion*.

deposits in the Argentine banking system were pulled from the country, much of it going to neighboring Uruguay. Although more than two-thirds of these deposits had returned by the end of 1995, Argentina also suffered a severe recession in 1995, with GDP declining by 4.4 percent (Inter-American Development Bank 1997).

Effects of the Asian Crisis. If the Mexican crisis showed that Latin American economies are susceptible to contagion effects, merited or otherwise, stemming from crises occurring elsewhere in the region, the Asian crisis of 1997–98 demonstrated that contagion from one crisis can threaten geographically distant economies in both developed and developing areas.

The impact of Asia's financial difficulties on Latin America was most visible through financial market contagion. As developments in Asia unfolded, capital flows into Latin America also slowed, declining sharply in late 1997 and 1998. After surpassing $100 billion in 1996 and 1997, private capital inflows to Latin America declined to $85 billion in 1998 and are forecast to diminish further in 1999, to $66 billion (Institute of International Finance 1999). Most of the flows in 1998 took the form of foreign direct investment as portfolio investment turned sharply negative. Stock indexes throughout the region lost ground during the fourth quarter of 1997 and much of 1998 with the Russian debt default in August of that year also contributing to the slump. (Chart 2 displays the performance of major Latin stock indexes from October 1997 through 1998.) Confidence in the Brazilian market during 1998 and early 1999 eroded to the point that capital flight compelled authorities to devalue and eventually float the *real* in January 1999.

Weakened currencies and recession in Asia during 1998 also affected Latin American trade balances. Chile in particular is heavily exposed to Asia with 34 percent of its 1997 exports destined to that region; Peru followed with 22.6 percent. Meanwhile, virtually all economies in the region depend significantly on primary exports—such as petroleum and copper—the prices of which experienced major declines in 1998, in no small part resulting from reduced Asian demand. (South Korea and Japan, for example, are among the world's largest consumers of copper.) Latin American export revenues actually declined slightly in 1998, from $254 billion to $248 billion, and the region's trade deficit increased from $31 billion in 1997 to $50 billion in 1998 (Institute of International Finance 1999). Partially as a result of the Asian crisis, Latin American regional GDP growth decelerated significantly in 1998, reaching only 2.0 percent after a 5.1 percent rate the year before. Early forecasts for 1999 suggest a mild recession for the year, with the regional economy contracting by 0.5 percent (Latin American Consensus Forecasts 1999).

Policy Responses to Financial Contagion

The effects of the Asian economic crisis have underscored Latin America's ongoing vulnerability to external shocks through both trade and financial channels. The crisis highlighted the region's limited export base and continued reliance on primary commodities as a source of foreign exchange, and the massive scale of capital inflows during the 1990s resulted in greater exposure of the region's economies to international financial volatility. These effects have

TABLE 2 Real GDP Growth for Selected Latin American Economies (1995–99)

	1995	1996	1997	1998	1999f
Argentina	-4.0	4.8	8.6	4.2	-3.2
Brazil	3.0	2.9	3.7	0.2	-0.5
Chile	10.6	7.4	7.1	3.4	-0.1
Colombia	5.8	2.1	3.1	0.6	-1.6
Ecuador	2.3	2.5	3.4	0.4	-4.3
Mexico	-6.2	5.2	7.0	4.8	3.0
Peru	7.3	2.5	7.2	0.7	2.9
Venezuela	3.4	-1.6	6.0	-0.7	-5.7

f = forecast

Source: Inter-American Development Bank data and Latin American Consensus Forecasts

jump-started further debates on appropriate economic policies for the region. Thus far, how have Latin American policymakers responded to the latest bout of external economic shocks?

Not surprisingly, the answer depends on each country's starting point. Economies that had already introduced comprehensive, market-oriented reforms prior to the crises—such as Chile, Mexico, and Peru—have not introduced substantially different economic policies. All of them have recently implemented some degree of fiscal tightening to account for the impact of weaker commodity prices on export earnings (Institute of International Finance 1999). In other words, authorities in these economies have pursued procyclical fiscal policies, tightening the policy regime during an economic downturn. Under normal conditions, one would expect countercyclical fiscal policy, but because global capital markets tend to question emerging markets' fiscal rectitude during contagion episodes, procyclical policies are often deemed necessary to restore market confidence (Hausmann and others 1996).

Other economies in the region, in particular Brazil, were more significantly affected by the Asian crisis because of their internal economic imbalances. In Brazil a nominal fiscal deficit that reached 6.1 percent of GDP in 1997 prompted market unease and subsequent capital flight (Central Bank of Brazil 1998). As a result, Brazil has begun a major fiscal adjustment process and has received substantial multilateral financial support for this effort. Ecuador and Venezuela, which as oil-exporting nations were especially vulnerable to declining commodity prices, are facing similar challenges.

In hindsight it appears that the economies that had carried out the most significant structural reforms earlier in the decade (or in the case of Chile, as far back as the 1970s and 1980s) were best able to insulate themselves from the global turmoil. Mexico and Peru are both expected to post moderate GDP growth in 1999 while economic activity in Chile is forecast to remain flat. Economies such as those of Brazil and Colombia, which have only partially implemented economic reforms, are expected to contract this year. Finally, nonreforming economies like Ecuador and Venezuela are forecast to experience sharply negative growth in 1999 (see Table 2) and thus still face serious policy challenges. Overall, countries that have undergone significant market-oriented reforms in the form of trade liberalization, fiscal adjustment, and privatization are posting relatively stronger performance than those that have not. (The only reformist nation in the region to suffer severe contagion effects as a result of the Mexican crisis of 1994–95 and the Asian crisis of 1997–98 was Argentina, which suffered from a perceived dual vulnerability: its fixed exchange rate regime and its high trade exposure to Brazil.)

Conclusion

During most of the economic crises of the post–World War II period, governments in Latin America responded by introducing new economic policies representing a significant departure from the prior policy path. Post–Depression era declines in the terms of trade for primary goods served as the impulse for import substitution policies in the immediate postwar period and through the 1960s. The availability of international liquidity during the 1970s, in conjunction with the oil shocks and the subsequent international recession, helped spark the debt crisis of the 1980s. This crisis in turn led to the adoption of comprehensive market-oriented reforms in several Latin American economies during the late 1980s and the 1990s.

The adoption of market-oriented reforms did not render the region invulnerable from external shocks, as

the impact of the Asian crisis on Latin America amply demonstrates. But unlike the earlier crisis episodes, which prompted a sea change in the economic policy regime of several economies, the recent external shocks have led most governments to deepen, not depart from, the overall market-oriented framework. Although it took the region more than a decade to recover from the debt crisis of the 1980s, the recovery time from the recent crises is expected to be briefer: most forecasts suggest moderate economic growth for Latin America in 2000, following a regional downturn in 1999. Latin America is thus proving more resilient under the market framework of the 1990s than under the state-led economic policies of earlier decades.

REFERENCES

BIRDSALL, NANCY, MICHAEL GAVIN, AND RICARDO HAUSMANN. 1998. "Getting the Lessons Right: A View from the Inter-American Development Bank." In *Mexico 1994: Anatomy of an Emerging Markets Crisis*. Edited by Sebastian Edwards and Moises Naím. Washington, D.C.: Carnegie Endowment for International Peace.

BIRDSALL, NANCY, CAROL GRAHAM, AND RICHARD SABOT. 1998. "Virtuous Circles in Latin America's Second Stage of Reforms." In *Beyond Tradeoffs: Market Reform and Equitable Growth in Latin America*. Edited by N. Birdsall, C. Graham, and R. Sabot. Washington, D.C.: Inter-American Development Bank and Brookings Institution Press.

BIRDSALL, NANCY, AND CARLOS LOZADA. 1998. "Shocks Externos en Economías Vulnerables: Una Reconsideración de Prebisch." In *CEPAL Review*. Economic Commission for Latin America and the Caribbean. Santiago: United Nations, October.

CENTRAL BANK OF BRAZIL. 1998. *Boletim do Banco Central do Brasil*, July.

CHANG, ROBERTO. 1999. "Understanding Recent Crises in Emerging Markets." Federal Reserve Bank of Atlanta *Economic Review* 84 (Second Quarter): 6–16.

EDWARDS, SEBASTIAN. 1993. *Latin America and the Caribbean: A Decade after the Debt Crisis*. Washington: World Bank.

HAUSMANN, RICARDO, ERNESTO TALVI, MICHAEL GAVIN, AND ROBERTO PEROTTI. 1996. "Managing Fiscal Policy in Latin America and the Caribbean: Volatility, Procyclicality, and Limited Creditworthiness." Inter-American Development Bank, Office of the Chief Economist, Working Paper 326, March.

INTER-AMERICAN DEVELOPMENT BANK. 1985. *Economic and Social Progress in Latin America*. Washington, D.C.: Inter-American Development Bank.

——— . 1992. *Economic and Social Progress in Latin America*. Washington, D.C.: Inter-American Development Bank and Johns Hopkins University Press.

———. 1996. *Economic and Social Progress in Latin America*. Washington, D.C.: Inter-American Development Bank and Johns Hopkins University Press.

———. 1997. *Economic and Social Progress in Latin America*. Washington, D.C.: Inter-American Development Bank and John Hopkins University Press.

———. 1998. *1997 Annual Report*. Washington, D.C.: Inter-American Development Bank.

INTERNATIONAL MONETARY FUND. 1995. *International Capital Markets: Developments, Prospects, and Policy Issues*. Washington, D.C.: IMF, July.

INSTITUTE OF INTERNATIONAL FINANCE, INC. 1999. "Capital Flows to Emerging Market Economies." Washington, D.C.: Institute of International Finance, Inc., April 25.

KUCZYNSKI, P. 1988. *Latin American Deb*t. Baltimore: Johns Hopkins University Press.

LATIN AMERICAN CONSENSUS FORECASTS. 1999. June 21.

PREBISCH, RAUL. 1963. *Hacia Una Dinámica del Desarrollo Latinoamericano*, Suplemento de Comercio Exterior, Banco Nacional de Comercio Exterior, S.A, Mexico.

ROSENTHAL, GERT. 1996. "Development Thinking and Policies: The Way Ahead." In *CEPAL Review,* no. 60. Santiago: United Nations.

WILKIE, JAMES, ed. 1995. *Statistical Abstract of Latin America*. Los Angeles: UCLA Latin American Center Publications.

WORLD BANK. 1991. *World Development Report*. Washington, D.C.: Oxford University Press.

———. 1997. *Private Capital Flows to Developing Countries: The Road to Financial Integration*. Washington, D.C.: Oxford University Press.