Federal
Reserve
Bank of
Cleveland

Economic REVIEW

1988 QUARTER 3

ECONOMIC REVIEW

1988 Quarter 3
Vol. 24, No. 3

Rules Versus Discretion: Making a Monetary Rule Operational 2
by John B. Carlson

The question of whether monetary policy should be conducted by rules known in advance to all or by policymaker discretion is an enduring debate. This paper traces the evolution of rule advocacy from the time of the Federal Reserve Act and examines the problems associated with making a rule operational. The author concludes that some degree of discretion may be necessary and that it is useful to analyze policies on the continuum between a pure rule and pure discretion. Recent changes in the operating procedure illustrate how actual policy has varied within this continuum.

Actual Competition, Potential Competition, and Bank Profitability in Rural Markets 14
by Gary Whalen

Despite a great deal of previous empirical work, disagreement remains about the relationship between market concentration and bank performance. Further, the performance impacts of potential competition remain virtually unexplored. In this study, the author examines the relationship between bank profitability and both actual and potential competition. Risk is explicitly incorporated into the analysis, and possible simultaneity is also explored. The results suggest that non-MSA markets are contestable or that potential competition has a significant influence on banks' performance.

Getting the Noise Out: Filtering Early GNP Estimates 24
by John Scadding

The accuracy of the U.S. Department of Commerce's estimates of GNP is vital in order for policymakers and private agents to make informed economic decisions. The early, or provisional, GNP estimates for each quarter can be seen as rational forecasts of the final numbers. This article takes an alternative view: that the early GNP numbers are estimates of the final number, but estimates that are contaminated with error, or "noise." The author investigates a way to adjust the numbers to remove the error and to produce more accurate predictions of the final GNP growth estimates.

Comment 32

Economic Review is published quarterly by the Research Department of the Federal Reserve Bank of Cleveland. Copies of the *Review* are available through our Public Information Department, 216/579-2157.

Coordinating Economist:
Randall W. Eberts

Editor: William G. Murmann
Assistant Editor: Robin Ratliff
Design: Michael Galka
Typesetting: Liz Hanna

ISSN 0013-0281

Rules Versus Discretion: Making a Monetary Rule Operational

by John B. Carlson

John B. Carlson is an economist at the Federal Reserve Bank of Cleveland. The author would like to thank Charles Carlstrom, Edward Gambier, William Gavin, James Hoehn, Mark Sniderman, and E.J. Stevens for their helpful comments.

Introduction

The rules-versus-discretion debate is the most enduring, if not the most central, issue in monetary policy. It concerns whether monetary policy should be conducted by rules known in advance to all or by policymaker discretion.

For many years, the case for a monetary rule was associated with a particular proposal by Milton Friedman (1959). Building on a tradition initiated by Henry Simons (1936), Friedman introduced the idea that the effects of monetary policy were uncertain, occurring with long and variable lags. In short, he argued that discretionary management of the money supply in the face of such uncertainty actually amplified economic fluctuations. Hence, Friedman argued for a constant-money-growth rule.

The case for rules has changed fundamentally since an important paper by Kydland and Prescott (1977). They show that precommitment to a rule could have beneficial effects that discretionary policies cannot. Unlike Friedman's argument, the Kydland-Prescott case was not specific to any one view of the world, but could be applied to a very general class of models. In principle, one cannot deny that a policy rule can have potentially stabilizing effects.

The example of Kydland and Prescott, however, trivialized an important concern of policymakers: how to account for uncertainty in the link between policy instruments and ultimate objectives. Once one allows for uncertainty, there is a potential role for flexibility to deal with variability in the links. To the extent that some variation is systematic and can be predicted, it is possible to incorporate feedback into a rule. However, some contingencies cannot be foreseen. When such events are potentially destabilizing, discretion may not be ruled out a priori.

This suggests that it is reasonable to consider the idea of rules with discretion. Fischer (1988) has concluded that the dichotomy between rules and discretion should be seen as a continuum, in which the extent of the monetary authority is determined by the immediacy of the link between its actions and the attainment of the objectives.

The actual practice of monetary policy can be viewed as a point on the continuum. Moreover, the rise of monetary targeting in the 1970s, which led to alternative operating procedures with differing degrees of commitment, illustrates that the degree of commitment to any rule can vary over time. Changes in the degree of commitment are best understood when one confronts the difficulties in making rules operational.

This paper reviews the historical development of the rules-versus-discretion debate and examines the problems associated with making rules operational. Section I traces the evolution of rule advocacy from the time of the Federal Reserve Act. Section II describes the actual operating procedures from the early 1970s to the present. The operational problems facing rule advocates are highlighted in Section III, and Section IV discusses how two recently proposed rules address the operational problems. Section V offers some concluding comments.

I. Rule Advocacy in the United States After the Federal Reserve

The original Senate bill to create the Federal Reserve System in 1913 contained a provision that the system should *promote a stable price level.* This provision was stricken by the House Committee on Banking and Currency and was not included in the original Federal Reserve Act, reflecting the dominant influence of the real bills doctrine at that time. By the late 1920s, however, several bills had been proposed to amend the Federal Reserve Act explicitly to include a provision for price stability.[1] Advocates of these bills essentially sought to legislate a rule establishing the primacy of the price-level objective.

These efforts culminated in the Strong Hearings, held by the House Banking Committee in 1926-1927.[2] The hearings initially considered a bill by Representative James G. Strong including a provision that "all the powers of the Federal Reserve System should be used for promoting a stable price level." Specifically, Congressman Strong did not want the Federal Reserve to have the discretion to vary the price level for the pursuit of any other objective.

While the bill instructed that the Federal Reserve's discount-rate policy was to be determined with "the view of promoting price stability," no formula was specified. Thus, there was a certain vagueness about how the rule would be implemented.[3] It left open the role for discretion in determining how much the discount rate should be altered when the price level deviated from its objective. A subsequent version of the bill was even more ambiguous about the objective of price stability. Eventually, Congressional interest in establishing the primacy of the objective of price stability faded.

The Simons Tradition

In a widely celebrated article of 1936, Henry Simons initiated a case for rules that was to become known as the Chicago view. Specifically, Simons contrasted two sharply distinct ways to conduct monetary policy: one, to assign in advance specific responsibilities to a monetary authority to be carried out in accordance with well-defined operational rules; the other, to specify a general goal while allowing the monetary authority wide discretionary powers to achieve the goal. The essential distinction is that the first regime defines the authority's objective in terms of the means, while the second defines the objective in terms of the ends.

Simons argued for rules in terms of means. His case was predicated on liberal (19th-century sense) principles. "The liberal creed demands organization of our economic life largely through individual participation in a game *with definite rules.* It calls upon the state to provide a stable framework of rules within which enterprise and competition may effectively control and direct the production and distribution of goods." (Simons [1936], p. 1)

The essential notion is that government is necessary for establishing laws that would define the rules for a "game" in which competitive free enterprise could flourish, but that government should not be a player in the game. The idea that government would manage the currency to manipulate aggregate economic outcomes meant that government would be a player and thus violated the liberal creed.

An ideal rule according to Simons would be one that fixed the quantity of the money supply. He did not believe, however, that such a rule could be made operational without radical reform of the financial structure. Essentially, he believed that an unregulated financial sector was a source of great instability in money demand. This instability was reflected in the perverse behavior of velocity which, he argued, necessitated a role for discretionary actions. Simons

■ 1 For a thorough review of the debate, see Fisher (1934). It should be noted here that a provision for purchasing power was eventually incorporated in the Employment Act of 1946. However, the price-stability goal was not included as the primary objective as most advocates of price stability in the 1920s had sought.

■ 2 For an excellent discussion of the background and events surrounding the Strong hearings, see Hetzel (1985).

■ 3 Hetzel (1985) notes that Congressman Strong and his supporters wanted to institutionalize the policy of Governor Strong (no relation) of the New York Federal Reserve Bank, which they credited for the considerable price stability that existed after 1922.

therefore suggested a number of ideal reforms to reduce the variability of velocity to levels conducive to successful implementation of a fixed-money-supply rule. That is, government would need to redefine the rules of the game to avoid having to manage the money supply.

One proposed reform was the elimination of fractional-reserve banking. By requiring 100-percent reserves on all demand deposits, Simons sought to reduce greatly the threat of bank runs and the consequent effects on hoarding money (velocity changes). Such a reform would also give the monetary authority direct control over the total money supply by making it equivalent to the monetary base.

Simons recognized, however, that fixing the supply of deposits might merely serve to encourage the creation of effective money substitutes that would also affect velocity. Thus, another "ideal" (but even more radical) reform would be to prohibit fixed-money contracts. Restricting claims to residual equity or common-stock form would essentially drive a wedge between money and other assets and would tend to minimize the variability of velocity. In sum, Simons believed that a monetary rule in terms of means could be made operational only under a highly regulated financial system.

Simons was not naive about the kind of assent that could be gained for such radical reforms in modern democratic societies. He thought that adoption of an appropriate framework could be implemented only after decades of "gradual and systematic reordering of financial practices." Ironically, liberal principles also seem to support the notion that financial institutions should be largely unregulated and free to offer any instruments they choose. Indeed, institutional reform has moved in the opposite direction of Simons's ideal.

Recognizing the practical difficulties of sharp changes in velocity and that his ideal reforms might be unattainable, Simons argued for a rule for price stability in the interim. Because this is a rule of ends rather than means, the operational procedures were not well defined. His basis for this practical solution was that it was the "least illiberal" of the alternatives he considered. Thus, he recognized that for immediate purposes a certain amount of discretionary latitude was necessary. While Simons may have misjudged society's willingness to adopt his ideal reforms (new rules), his liberal view of economic agents participating in a game was prescient about the future state of the debate.

The Simons tradition was subsequently modified and popularized by Milton Friedman (1948, 1959, 1969). Initially, Friedman offered detailed proposals much in the spirit of Simons. They included the 100-percent reserves reform applied to both time and savings deposits at banks.

Subsequently, however, Friedman changed tack, taking the position that the behavior of velocity, particularly the velocity of the M2 aggregate, was not so perverse in a relative sense, even under a fractional-reserve banking system. He argued that the discretionary actions of the Federal Reserve (albeit well-intentioned) were likely to be a more perverse source of economic instability. Thus, adherence to a constant-money-growth rule would lead to greater economic stability than would a rule with feedback, with or without discretion. In essence, Friedman maintained the idea that the monetary authority should not be a player in the game, but he eventually rejected the need for wholesale reform of the financial system.

Friedman's case for a constant-growth rule was based less on the liberal creed and more on pragmatism. His premise was that the economic impact of monetary policy occurs with a long and variable lag. Feedback, especially of the discretionary type, would have effects at the inappropriate time more often than not. Moreover, Friedman argued that political pressures and accountability problems under discretion are likely to exacerbate the problem.

While Friedman's case has intuitive appeal, it is difficult to justify in principle. Potentially stabilizing effects of policy feedback could be ruled out a priori only if money were the exclusive determinant of nominal GNP in the short run. If other identifiable factors also have significant explanatory power, then judicious use of feedback can, in principle, reduce the variability of nominal GNP, even if the coefficients on lagged money are stochastic. On the other hand, the stabilizing effects of policy feedback with parameter uncertainty are smaller than when parameters are nonstochastic (see Brainard [1967]).[4]

By eventually abandoning 100-percent reserves, Friedman also allowed a control problem: how to make a constant-growth rule operational for measures of inside money. Under 100-percent reserves there would be virtually no distinction between money and monetary base. Since Friedman also proposed closing the discount window, all money would essentially be outside money, and hence directly controlled by the Fed.

■ **4** When effects of monetary policy occur with a lag, there is a potential for instrument instability. The prospect of dynamic instability can be reduced with appropriate modifications to the objective function.

As advocates for constant money growth dropped the idea of 100-percent reserves, however, the issue of monetary control became relevant. When the measure of money is endogenous, the problem of making a constant-money-growth rule operational is far from trivial. Such was an important lesson of monetary targeting in the early 1980s. Perhaps recognizing this fact, advocates for money-growth rules now typically propose closing the discount window and adopting a constant-growth rule for the monetary base.

Arguments for a monetary rule in the Simons tradition remain highly controversial in principle. One cannot rule out the possibility that an intelligent policymaker could effectively take account of incomplete information when deciding optimal monetary policy. As Barro (1985) notes, "if the policymaker were also well-meaning, then there was no obvious defense for using a rule in order to bind his hands in advance." Moreover, Fischer argues, "At a formal level Friedman's analysis suffered from the logical weakness that discretion seemed to dominate rules: if a particular rule would stabilize the economy, then discretionary policymakers could always behave that way—and retain the flexibility to change the rule as needed."

Kydland and Prescott

The idea that discretion could always replicate a preferred policy rule seemed to provide a highly influential argument in which intelligent, well-meaning policymakers should not be bound by rules. However, in a widely recognized paper, Kydland and Prescott (1977) demonstrate a fallacy in this argument. It is now well understood that if economic outcomes depend on expectations about future policies, then credible precommitment to a rule could have favorable effects on the economic outcomes that discretionary policies cannot have.

Applications of the Kydland and Prescott result to monetary policy are often developed in familiar (and highly abstract) models of output and inflation.[5] These models assume that wage-setters and the monetary authority are engaged in a noncooperative game. In this game, wage-setters must specify the nominal wage rate in a contract (their play) *before* prices are determined (the policymaker's play). Firms' decisions to hire are made after prices are determined, so that the real wage is known. Since firms are assumed to be profit maximizers, the real wage determines the level of output for the economy.

An essential feature of the game is that by determining the price level, the policymaker's play determines the real wage and level of real output. Moreover, expectational errors of wage-setters determine the deviation of output from its full employment levels. Thus, the game yields the familiar output supply function

$$(1) \quad y = y^* + b(\pi - \pi^e),$$

where y and π are output and inflation, y^* is full employment output, and π^e is the expected inflation rate.

The policymaker is assumed to have a loss function quadratic in the deviations of inflation and output from target levels. Here, desired inflation is assumed to be zero.

$$(2) \quad L = a\pi^2 + (y - ky^*)^2$$

The target rate of output is assumed to be above the natural rate, that is, $k > 1$. One motivation for this assumption is that tax distortions and unemployment policy cause the natural rate to be too low from a social point of view. Alternatively, one might argue that the labor market is dominated by large unions (see Canzoneri [1985]). He assumes that the labor supply curve includes only union members and that wage-setters' behavior systematically excludes other workers. By contrast, the loss function includes all workers. Others have argued that equation (2) is not really a measure of social utility, but reflects the bias of policymakers to underestimate the natural rate of unemployment.

To illustrate the advantage of a rule, consider the case in which the policymaker has discretion in a one-period game. Because the policymaker chooses policy *after the wage-setters specify the wage rate,* the wage-setters know that the policymaker has the incentive to take the *expected inflation rate as given* and to induce higher employment with additional inflation, if possible. Given the known loss function, there is only one strategically rational expectation (that is, Nash solution) for inflation:

$$(3) \quad \pi^e = a^{-1}b(k - 1)y^*.$$

■ **5** The particular example presented here is the compact static model in Fischer (1988). The use of a *static* model to illustrate *dynamic* inconsistency has been criticized as inadequate. The basic concept, however, has been developed in the context of a dynamic model (see Roberds [1986]). Since it is the concept we want to convey here, the static model suffices.

Under this solution, the policymaker has no incentive to choose an inflation rate higher than expected. The gains from the additional output would be more than offset by the loss of the additional inflation. Note also that if the policymaker had an objective for the inflation rate less than the expected inflation rate *before* wage-setters acted, it would be inconsistent *afterward*. That a zero-inflation objective is not credible with discretion is an example of the problem of *time inconsistency*.

The value of the loss function evaluated at the solution is denoted as L_d and is given by

$$(4) \quad L_d = (k-1)^2 y^{*2} [1 + a^{-1} b^2].$$

If the policymaker could credibly precommit to a policy of zero inflation, that is, a dynamically consistent inflation objective, the loss function would be

$$(5) \quad L_p = (k-1)^2 y^{*2}.$$

Since $L_p < L_d$, precommitment to a zero-inflation objective affects expectations in a way that leads to a more favorable outcome than pure discretion would allow. Essentially, discretion buys nothing in terms of output, which is the same under both policies, but leads to an inflationary bias.

To be sure, the basic result of Kydland and Prescott demonstrates in a very precise way a benefit to precommitment to a policy rule. Although developed in a highly abstract model, the result has been widely influential in academic research. A major shortcoming of the analysis, however, is that it trivializes the control problem. Specifically, it presumes that the policymaker has a deterministic operating procedure that enables precise control of inflation. Once disturbances are introduced into the model, the precommitment solution does not necessarily dominate the discretion solution.

To analyze the control problem, Canzoneri considers a stochastic disturbance to money demand such that velocity follows a random walk. In his game, wage-setters cannot see the disturbance at the time they specify their wage, but the Federal Reserve has some forecast of money demand before it chooses its policy for money growth. If the Fed is left with some flexibility, it can accommodate the predictable component of the change in velocity. As Canzoneri notes, this practice benefits both wage-setters and society as a whole. Thus, the policy problem becomes one of trading off flexibility needed for stabilization with the constraint needed for eliminating the inflation bias.[6]

The discussion thus far has been in the context of a one-period framework. In reality, however, the central bank has a horizon that extends beyond one period. Indeed, this may explain why central banks are typically isolated from political pressures by design. It is now widely understood that in a multiperiod context, the Fed may be able to establish a reputation that serves the same purpose as a monetary rule. This possibility has been investigated by Barro and Gordon (1983a, 1983b). They find that under certain conditions, reputation-building can lead to a result that is superior to pure discretion, although not as good as precommitment to a rule.

Barro and Gordon assume, however, that wage-setters eventually have access to the same information as the Fed. Canzoneri shows that when the Fed has its own private forecast of money demand, it has an incentive to misrepresent its intentions.[7] He further demonstrates that no stable resolution of the credibility problem can rely on the Fed's own announcement of its forecast. When the Barro and Gordon model is modified to account for asymmetric information, the Fed cannot build sufficient credibility by simply running a noninflationary policy for a few periods.

Rogoff (1985) has shown that other solutions may mitigate the problem of dynamic inconsistency. One such solution is that society can benefit by choosing a "conservative" central banker—one that places a high cost on inflation. In the context of the simple model above, this means that the central bank places a high value on parameter a in its loss function. Equation (4) reveals that as a gets large, the value of the loss function diminishes, ultimately approaching the value of the precommitment solution given in equation (5).

Like Barro and Gordon, Rogoff assumed symmetric information. When the Fed has private information, it has the incentive to appear more conservative than it actually is; the wage-setters have no way of telling. The implication is that there could be periodic inflationary breakdowns followed by sustained periods where credibility builds and wage-setters learn the true intentions of their central bank. Unfortunately, Canzoneri shows that it is no simple matter to legislate incentive-compatible rules that would remedy the problem posed by private information.

■ **6** Fischer (1988) demonstrates in a formal model that when control error exists, the ordering of the loss functions under precommitment and discretion is ambiguous.

■ **7** If the money demand forecast were predicated on a stable model over time, it would be preferable for the Fed to commit to a contingent rule based on that model forecast. Thus, while the rule would allow flexibility, it would not admit discretion. Given the absence of evidence of stability in money demand, such a rule seems infeasible.

Rogoff also demonstrates that under certain conditions, intermediate targeting may also provide a reasonable solution to the problem of dynamic inconsistency. By providing the central bank with incentives to hit an intermediate target, it is possible to induce fewer inflationary wage bargains in the context of his model. While the Rogoff result demonstrates some a priori basis for intermediate targeting, his analysis abstracts from many problems the policymaker faces in practice. Nevertheless, the literature since 1977 suggests there is a reasonable basis for some precommitment—if not to a rule for all time—to some monetary policy on a continuum between a pure rule and pure discretion.

II. The Operating Strategy of the Federal Reserve

The operating strategy of the Federal Reserve can be viewed as a commitment to a policy on the continuum between a pure rule and pure discretion. The rule-like elements are embedded in the Fed's monetary targeting procedure. Monetary targets are not ends in themselves, but are intermediate variables between the instrument variables that the Fed directly controls, such as the federal funds rate or nonborrowed reserves, and ultimate goals, such as price stability and stable output growth. Thus, intermediate target variables must be closely linked to both ultimate objectives and instruments.

The use of intermediate targets has been criticized as redundant and inefficient from a control-theoretic perspective (see B. Friedman [1975]). These objections, however, are based on the assumption that policymakers have precise, reliable knowledge about the relationships between instruments and final objectives. In practice, policymakers see great uncertainty in these links and doubt that such relationships could be captured by econometric models accurately enough to be operationally useful (see Black [1987]). In contrast, intermediate target variables are seen as relatively more controllable than ultimate variables.

Moreover, policy decisions are made by majority rule. It is therefore difficult, if not impossible (Arrow's theorem) to obtain a consensus for adopting a particular social objective function, which is necessary under direct targeting of final objectives. Under an intermediate targeting strategy, the Fed does not need to specify numerical objectives for goal variables.

Intermediate targeting strategies can vary substantially in degree of flexibility or commitment. In principle, intermediate targets may or may not be designed to allow feedback. For example, a target could be specified for a five-year horizon without allowing for revisions, or for a three-month horizon to accommodate frequent adjustments based on new information. Also, the operating procedure used to control the target variable may or may not allow for a high degree of discretion. Thus, operating rules could be highly automatic with infrequent discretionary input or be judgmentally modified day-to-day, based on the latest information.

Actual practice of monetary targeting indicates that the degree of flexibility and discretion incorporated into the strategy is influenced by two key factors. The first is evidence concerning the stability of the relationship on which the strategy is based. If there is a broad consensus about the reliability of the relationship between the intermediate target and ultimate goals, then it is more likely that a central bank would be willing to commit to closer targeting of the variable with less feedback from other sources, whether discretionary or not. The other key factor is the central bank's credibility or reputation in containing inflationary expectations. If the central bank establishes its credibility by avoiding inflationary policies, then the public and Congress are generally more willing to accept a greater degree of discretion in strategy and tactics.

The interplay of these factors may well account for the increased reliance on monetary aggregates as intermediate targets during the early 1970s. Before the mid-1960s, there was scant evidence that discretion exercised by the Federal Reserve provided a substantive basis for inflationary expectations. Nominal interest rates were, on average, too low to indicate expectations of rising inflation. The public apparently believed that the Fed would "take the punchbowl away just as the party got going," a perception consistent with Rogoff's notion of a conservative central bank. Although the Federal Reserve had intermediate targets for interest rates—a strategy that is now widely viewed as potentially defective for avoiding inflation—the Fed seemed to use its discretion judiciously in avoiding inflation and hence in assuaging public doubt about the efficacy of its operating strategy.

By the early 1970s, however, a basis for doubt was beginning to emerge, as inflation had accelerated to new and persistently high levels. Over that decade the Fed gradually strengthened its reliance on monetary aggregates as a source of information about its ultimate objectives.

While the process was initially internal only, the Fed began to announce publicly its desired annual growth ranges for selected monetary aggregates in response to a Congressional resolution in 1975. Evidence in the early 1970s convinced many that the relationship between money and nominal GNP—as summarized by velocity—was sufficiently reliable to choose monetary targets over annual, or even longer, horizons. Also, the parallel rise in the price level offered simple but persuasive evidence that inflation could be slowed by slowing growth of the monetary aggregates. In 1979, the Fed adopted a strategy for disinflation by gradually reducing the rate of money growth from year to year.

The strategy was coupled with an automatic feedback rule to enhance monetary control and demonstrate a commitment to the strategy. Over most of the 1970s, the Fed used the federal funds rate—the interest rate banks charge one another on overnight loans of reserves—as its operating target for controlling money growth. Specifically, it sought to influence the quantity of money the public demanded by altering the opportunity cost of money. For example, if money growth was too rapid, it attempted to raise the federal funds rate, and thereby raise other short-term rates.

The higher rates were expected to slow money growth by inducing the public to shift from monetary assets to other financial assets. Over longer horizons, higher interest rates might also be expected to slow spending growth and hence the transactions demand for money. In practice, however, there is always substantial pressure for the Fed to minimize interest-rate movements, particularly interest-rate increases. For this reason and others, the Federal Reserve did not respond sufficiently promptly or intensively to keep monetary growth from accelerating in the 1970s.

By late 1979, the inflation rate had accelerated to double-digit levels. Financial markets, especially foreign markets, began reacting strongly to the inflationary developments. The dollar was falling rapidly as foreign investors appeared to doubt the Fed's resolve to contain inflation. In response to the evident inflationary pressures, the Federal Reserve adopted a new set of tactics "as a sign of its commitment to longer-run restraint on money growth" (Lindsey [1984], p. 12). These tactics in effect eliminated a substantial degree of discretion that the Fed had used to smooth short-term interest-rate movements.

The new procedures sought to control money growth by maintaining a short-run target path for nonborrowed reserves. As Lindsey describes, "holding to a nonborrowed reserves path essentially introduces in the short run an upward sloping money supply curve on interest rate and money space" (p. 12). In effect, the nonborrowed reserves target created an *automatic* self-correcting mechanism that would partially resist all deviations of money from target. If money growth in a given week moved above target, the prespecified level of nonborrowed reserves virtually assured that the federal funds rate would move upward. In sum, the Federal Reserve gave up its discretion to minimize federal funds rate movements to assure financial markets of its commitment to the disinflation strategy.

While the new procedure involved substantial commitment at the tactics level, it permitted significant discretionary feedback at the strategy level. Under the strategy, the FOMC was free to change its short-term monetary target to take account of new information—a practice that led to significant deviations of money from announced annual targets. Such discretionary feedback was deemed necessary as evidence mounted that the velocity of money was not as reliable as expected.

It was well understood at the time that deregulation in financial markets, changes in transactions technology, and disinflation were having a substantial impact on individual portfolios and hence on the velocity of money. While such factors could account for the target misses in a qualitative sense, policymakers lacked means to predict the impact on money growth in order to specify reliable target values. By August 1982 the evidence was compelling that the behavior of velocity had been altered in some permanent way. Because time was needed to identify the new patterns of velocity behavior, attempts to control monetary aggregates closely appeared futile.

Consequently, the Fed abandoned its operating procedure and hence its commitment to a fixed path of nonborrowed reserves in the short run. It de-emphasized the role of M1 and adopted a more flexible operating strategy. Since the fall of 1982, the Fed's operating target has been the aggregate level of seasonal plus adjustment borrowings at the discount window. Under this procedure, the FOMC specifies a short-term objective for this variable at each of its regularly scheduled meetings (at approximately five- to six-week intervals).

Unlike with the nonborrowed reserves operating target, the current procedure does not produce automatic self-correcting federal funds rate responses to resist divergences of money from its long-run path. Substantial changes in the federal funds rate are largely a consequence of judgmental adjustments to the borrowings target. Thus, the Fed has regained much of the leeway to smooth short-term interest rate changes that it had prior to 1979.

It is important to note that by the end of 1982 the disinflation process had become credible to most of the public. Financial markets, particularly those for fixed-income securities, reacted favorably to the procedural change. Long-term interest rates continued to decline substantially after the Fed announced abandonment of the nonborrowed-reserves procedure. Moreover, over the long term, wage demands moderated to pre-1970s levels and have been persistently moderate to this day.[8] Such would seem strong evidence that wage-setters haven't suspected the Fed of "cheating" on its goal of reducing and maintaining lower inflation.

The evolutionary cycle of the Federal Reserve's operating procedure provides a useful illustration of how the degree of discretion has varied in response both to evidence concerning the reliability of the money-income relationship and to the reputation of the Fed. As the Fed's credibility on inflation appeared to wane in the 1970s, it adopted procedures that increased reliance on monetary aggregates as intermediate targets and limited its discretion to smooth interest rates. As evidence suggested a breakdown in the behavior of velocity, the degree of commitment to monetary control diminished to allow the necessary operational flexibility. By that time the Fed's commitment to maintaining lower rates of inflation had become credible. While the actual strategy can be characterized as a monetary rule with varying degrees of discretion, it never incorporated the degree of commitment that most monetarists had hoped for—one that would have not altered monetary targets at all.

III. Problems with Making Rules Operational

The review of the Federal Reserve's actual operating strategy also serves to highlight a number of potential problems with making rules operational. Poole (1988), a longtime monetary rule advocate, recently concludes that "there is a serious and probably insurmountable problem to designing a predetermined money growth path to reduce inflation." Essentially, he argues that it is not possible to reliably quantify the effects of disinflation on money demand and, hence, on velocity.[9] Thus, managed money is unavoidable during the transition to lower inflation. While Poole accepts the eventual efficacy of a constant-growth rule, he believes there is no formula to determine when the discretionary mode should terminate. Presumably, it would only be after inflation has been eliminated.

Even if the transition to lower inflation were no longer operationally relevant, the experience of the early 1980s makes it clear that money demand and velocity have also been independently affected by regulatory change and by developments in transactions technology. McCallum (1987) has recently argued that a rule should not rely on the presumed absence of the effects of such changes. This principle of rule design precludes simple, fixed rules like the constant growth rate of money (or monetary base). Operational feasibility demands that a monetary rule should at least be flexible enough to accommodate the effects of such changes on velocity.

Recognizing a need for some form of flexibility, some pure-rule advocates now propose nondiscretionary feedback rules. Nondiscretionary feedback requires specification of a formula linking goal (or target) variables to policy instruments. The formula presumes the existence of some reasonably stable and hence reliable model, that is, one that characterizes sufficiently well the relationship between instruments and objectives.

The absence of a consensus in macroeconomics about an appropriate model poses a serious obstacle for gaining assent for any *particular* feedback rule in practice. While most economists adopt *a* perspective, few seem willing to accept the notion that a particular (especially simple) characterization of the economy would be sufficiently reliable for long periods. Even among rule advocates sharing a common perspective, there are likely to be subtle differences about the formula specification that may splinter support for a given rule.

This problem of model uncertainty is compounded by the important demonstration by Lucas (1976) that "structural" models are in general not invariant to the way in which policy is implemented. Since this critique, there has been no widely accepted means of evaluating operationally concrete policy proposals.[10] While many large-scale econometric models have met the market test, few economists seem convinced by policy evaluations based on particular econometric models.

■ **8** For evidence concerning moderation in compensation demands, see Groshen (1988).

■ **9** This point is an example of a more general result of Lucas (1976), which is discussed below.

■ **10** Advocates of rules sometimes argue that if a nondiscretionary rule were to be implemented, relationships would stabilize, leading to more favorable outcomes than suggested by simulations based on historical relationships. While this purely a priori theoretical argument is consistent, it does not appear to be convincing to most economists.

Without a consensus about how monetary policy affects aggregate economic outcomes, it is not compelling to argue that expectations of economic agents (for example, wage-setters) are based on any one model of the economy. Any given rule could possibly be perceived as unsustainable by a sufficient number of agents such that the rule would not be credible in an aggregate sense. If agents believed the rule was unsustainable, the game between agents and policymakers would become extremely complicated, with no apparent solution. Thus, it would not be clear that commitment to a rule would be beneficial. It would seem useful that a rule advocate demonstrate that favorable consequences of a proposed rule would be robust to alternative models of the economy.

IV. Two Recently Proposed Rules

Two recently proposed rules by McCallum (1987, 1988) and Hall (1984) illustrate how the debate over rules versus discretion has evolved to a more operationally concrete level. Both authors appeal to the result of Kydland and Prescott as a justification for implementing their rules. Both also recognize a need for flexibility and address operational problems. In sharp contrast, however, is the way they incorporate flexibility.

McCallum proposes a nondiscretionary feedback rule for nominal income using the monetary base as the instrument. The target path of nominal income is fixed and grows at a prespecified rate of 3 percent per year. The feedback formula is

$$(6) \quad \Delta b_t = 0.00739 - (1/16)[v_{t-1} - v_{t-17}] + \lambda(x_t^* - x_{t-1}),$$

where b_t = log of monetary base (for period t), v_t = log of base velocity, x_t = log of nominal GNP, and x_t^* = target path for nominal income.

The constant term 0.00739 is simply a 3 percent annual growth rate translated into quarterly logarithmic units. The second term subtracts the average growth rate of velocity, approximated by the average difference in the logarithm of velocity over the previous four years. This term can be thought of as a simple time-series estimate of trend velocity growth. The third term specifies how policy is to respond to deviations of nominal income from its target path.

The moving average of velocity growth is a simple statistical filter designed to detect permanent changes in velocity growth. As such, it provides a mechanism to maintain a long-term correspondence between the current base growth path and the long-term nominal objective to account for changes in transactions technology. Given the length of the moving-average period (four years) and the absence of any systematic feedback from interest rates, however, the rule provides virtually no adjustment in response to the current state of the business cycle or to financial conditions.[11]

The third term provides feedback to assure that nominal income ultimately returns to its trend path. The choice of parameter λ incorporates some degree of flexibility to deal with the potential problem of instrument instability. This problem arises when effects of policy occur over time as they do in actual economies, particularly those with sticky prices. Large responses to maintain a target path in the near term could lead to longer-term effects in the opposite direction, requiring even greater offsetting policy responses in later periods. This sequence would be unstable if responses and effects were to become ever increasing. The value of λ (presumably less than zero) should be chosen to minimize the potential for this dynamic instability, under the constraint that it be sufficiently large to provide adequate responsiveness of base growth to target misses. McCallum suggests that a value of 0.25 appears to be somewhat robust for this objective over alternative models of the economy.

If velocity growth were constant, and if nominal GNP were on its target path for a sustained period, the policy prescribed by McCallum's rule would be the same as a 3 percent growth rule for the monetary base. Thus, McCallum's rule is essentially a generalization of the constant-money-growth rule. Because it is more general, it allows for flexibility to deal with some of the problems of making monetarist rules operational.

Moreover, McCallum claims that because the monetary base is "controllable," the rule can be accomplished with no operational discretion.[12]

■ **11** Recent evidence suggests that velocity has become increasingly interest-sensitive in the 1980s. To the extent that systematic effects of interest rates could be reliably estimated, additional flexibility could be introduced into the rule as feedback to compensate for short-run variability in velocity. McCallum expresses doubt, however, that economists know enough to base policy on any one short-run empirical model. In this sense he defends, if only indirectly, the monetarist dictum of Friedman, in which monetary policy affects the economy with long and variable lags.

■ **12** Under current institutional arrangements, the total monetary base can be controlled only indirectly, working through effects of changes in interest rates on the demand for base components. Advocates of base targeting often call for institutional reforms—such as exactly contemporaneous reserve accounting and closure of the discount window—to enable direct control of the base. Alternatively, McCallum's rule can be applied to the nonborrowed base, which is directly controllable under existing institutions.

In this sense McCallum's proposal is a flexible version of a rule for means. The flexibility is extremely limited, however, involving only feedback from simple statistical models to maintain long-run relationships. No role is given to structural models that might allow feedback for short-term economic stabilization. Such a rule shows little faith in macroeconomic models or in discretionary decisions of the Fed.

Some rule advocates, on the other hand, propose a much greater role for economic models and judgment of the Fed. An example is an ends-oriented rule advanced by Hall (1984). Under Hall's strategy, the Federal Reserve is instructed to stabilize the price level around a constant long-run average value. To make this strategy elastic in the short run, Hall proposes giving the Fed some prespecified leeway in achieving the target depending on the amount of unemployment. The permissible deviation of the actual price level, p, from its target, p^*, is defined by the simple numerical rule linking it to the deviation of the unemployment rate, u, from its normal rate, presumed to be 6 percent:

$$(7) \quad 100(p - p^*)/p^* = A(u - 6).$$

The coefficient A is to be specified by the Federal Reserve. Based on simulations, Hall tentatively recommends that it equal eight.

Specifically, this relationship is to be imposed as a constraint on policy instrument settings. In formal terms: "Monetary policy is on track when the deviation of the price level from its constant target is eight times the deviation of unemployment times its normal level [*presumed to be 6 percent*]. Policy is too tight if the price deviation is less than eight times the employment deviation; it is too expansionary when the price deviation is more than eight times the employment deviation. The elasticity of 8 in this statement is a matter for policymakers to choose." (Hall [1984], p. 140)

Policy formulation under this approach would be prospective. Thus, the Fed would need to employ a model that links instrument variables to the price level and to the unemployment rate over the criterion period.[13] It would be free to use whatever model and instruments it chooses. Instrument settings would be determined by an iterative process. To begin, an initial forecast for the unemployment rate and price level would be compared against the rule formula to be judged for appropriateness—for example, too tight, too easy, or on track. This process would thereby determine the direction in which instrument settings should be changed, if necessary. A second round of forecasts would then be obtained and compared. The process would continue until the instrument settings yielded price-level and unemployment forecasts consistent with the rule.

To impose discipline, Hall would require the Fed to be explicit about its forecasts, defending them publicly at the semiannual Congressional review and in comparison with private forecasts. Hall argues that forecasting errors of good private forecasters would provide a sufficiently reliable standard to maintain unbiased outcomes. If the Fed's forecasts were consistently different from reputable private forecasts, and if the outside forecasts were more often correct, then the Fed would be under public pressure to modify its way of setting policy instruments. For Hall, the problem with discretion lies not with the use of faulty econometric models but with the absence of a commitment to an explicit rule for the price level.

Both Hall and McCallum employ small empirical models to generate simulations under their rules. McCallum uses a variety of models based on competing views to examine the robustness of his rule's performance. His simulations suggest that his rule would have produced a root mean square error (RMSE) of nominal income of around 2 percent from 1954 to 1985. This is approximately one-third the RMSE of actual GNP around its trend over the same period. He concludes that his rule would have worked relatively well in the United States.

To address the criticism that his simulations are subject to the Lucas critique, McCallum notes that his rule relates nominal demand to nominal policy instruments. He argues that the sensitivity of parameters to policy regime changes is likely to be quantitatively less important for such rules than for rules that relate real to nominal variables, for example, based on Phillips curve models. Hall's simulations, on the other hand, are based on the presumption that there is a reliable (policy invariant) relationship between the *variability* of the inflation rate and the *variability* of the price level.[14] His simulation results suggest that *both* price level variability and unemployment variability would have been less than actually experienced from 1952 to 1983 under the elastic-price rule.

■ **13** Based on the assumption that monetary policy affects the unemployment rate reliably only after a yearlong lag, Hall argues that the criterion period should be the forecast horizon for the year beginning six months ahead.

■ **14** The analysis of policy in terms of the *variability* of unemployment and price level was developed by Taylor (1980, 1981). It is important to note that there is no implied trade-off in this model between the inflation rate and *trend* output growth.

While the results presented under both rules appear favorable, few analysts seem convinced by small-model simulations. Experience with large-scale econometric models, for example, suggests that interest rates would vary sharply under McCallum's rule. His models, which do not allow for interest-rate interactions, cannot account for the economic consequences of such interest-rate variation. Fischer (1988) argues that the natural vehicles for studying policy rules are the large-scale econometric models, many of which have met the market test. Nevertheless, he notes that it would be difficult to justify legislating any nondiscretionary rule given the variety and inadequacies of existing models. On the other hand, existing models may be no more reliable for discretionary decisions, particularly when policymakers may use them selectively to support their own prior beliefs.

V. Some Concluding Comments

The success of the U.S. disinflation strategy early in this decade helped reestablish the Federal Reserve's credibility as an inflation fighter. Much of the reputational capital surely persists today. Recently, however, some analysts have questioned whether the current strategy is adequate to extend and maintain the progress against inflation (see Black [1987]).

A key concern is that the strategy may lack sufficient institutional discipline to assure that short-term objectives—such as interest-rate smoothing—do not interfere with the achievement of longer-term price stability. This fear has led to a renewed interest in alternative strategies that are closer to a pure rule on the continuum between a pure rule and pure discretion.

Ideally, a policy strategy should perform adequately well under alternative views about aggregate economic relationships so that sufficient numbers of agents believe that the rule could be credibly implemented. Rule advocates might well follow the example of McCallum and examine the robustness of their rule's performance, simulating with alternative models of the economy. The choice of criteria for "adequate performance" is of course a difficult and controversial matter. We conclude here, as does Fischer (1988), that the discussion of alternative policies is too important to be suppressed by the econometric evaluation critique.

References

Barro, Robert J. "Recent Developments in the Theory of Rules Versus Discretion." *Conference Papers: Supplement to the Economic Journal.* 96(1985): 23-37.

_______, **and David B. Gordon.** "A Positive Theory of Monetary Policy in a Natural Rate Model." *Journal of Political Economy.* 91(1983a): 598-610.

_______, **and David B. Gordon.** "Rules, Discretion, and Reputation in a Model of Monetary Policy." *Journal of Monetary Economics.* 12(1983b): 101-21.

Black, Robert P. "The Fed's Anti-inflationary Strategy: Is It Adequate?" Federal Reserve Bank of Richmond. *Economic Review* (September/October 1987): 1-9.

Brainard, William. "Uncertainty and the Effectiveness of Policy." *American Economic Review Papers and Proceedings.* 57(May 1967): 411-25.

Canzoneri, Matthew B. "Monetary Policy Games and the Role of Private Information." *American Economic Review.* 75(December 1985): 1056-70.

Fischer, Stanley. "Rules Versus Discretion in Monetary Policy." National Bureau of Economic Research Working Paper No. 2518 (February 1988).

Fisher, Irving. *Stable Money: A History of the Movement.* New York: Adelphi Company, 1934.

Friedman, Benjamin M. "Targets, Instruments, and Indicators of Monetary Policy." *Journal of Monetary Economics.* 1(October 1975): 443-73.

Friedman, Milton. "The Lag in Effect of Monetary Policy." *Journal of Political Economy.* 69(October 1961): 447-66. Reprinted in *The Optimum Quantity of Money and Other Essays.* Chicago: Adeline Publishing Company, 1969: 237-60.

_______. *A Program for Monetary Stability.* New York: Fordham University Press, 1959.

_______. "A Monetary and Fiscal Framework for Economic Stability." *American Economic Review.* 38(June 1948): 245-64. Reprinted in *Essays in Positive Economics.* Chicago: The University of Chicago Press, 1953.

Groshen, Erica L. "What's Happening to Labor Compensation?" Federal Reserve Bank of Cleveland. *Economic Commentary,* May 15, 1988.

Hall, Robert E. "Monetary Strategy with an Elastic Price Standard." *Price Stability and Public Policy.* A symposium sponsored by the Federal Reserve Bank of Kansas City, August 1984: 137-59.

Hetzel, Robert L. "The Rules Versus Discretion Debate Over Monetary Policy in the 1920s." Federal Reserve Bank of Richmond. *Economic Review* (November/December 1985): 3-14.

Kydland, Finn E., and Edward C. Prescott. "Rules Rather than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy.* 85(June 1977): 473-91.

Lindsey, David E. "The Monetary Regime of the Federal Reserve System." Conference on Alternative Monetary Regimes sponsored by Ellis L. Phillips Foundation and Dartmouth College (August 1984).

Lucas, Robert E. "Econometric Policy Evaluation: A Critique." *Journal of Monetary Economics.* Supplementary Series 1(1976): 19-46.

McCallum, Bennett T. "Robustness Properties of a Rule for Monetary Policy." *Carnegie-Rochester Conference Series on Public Policy.* Carnegie Mellon University and National Bureau of Economic Research (February 1988).

_______ . "The Case for Rules in the Conduct of Monetary Policy: A Concrete Example." Federal Reserve Bank of Richmond. *Economic Review* (September/October 1987): 10-18.

Poole, William. "Monetary Policy Lessons of Recent Inflation and Disinflation." *The Journal of Economic Perspectives.* 2(Summer 1988): 73-100.

Roberds, William. "Models of Policy Under Stochastic Replanning." Federal Reserve Bank of Minneapolis Research Department Staff Report 104 (March 1986).

Rogoff, Kenneth. "The Optimal Degree of Commitment to an Intermediate Monetary Target." *Quarterly Journal of Economics.* C4(November 1985): 1169-89.

Simons, Henry C. "Rules Versus Authorities in Monetary Policy." *Journal of Political Economy.* 44(February 1936): 1-30.

Taylor, John B. "Stabilization, Accommodation, and Monetary Rules." *American Economic Review Papers and Proceedings.* 71(May 1981): 145-49.

_______ . "Output and Price Stability: An International Comparison." *Journal of Economic Dynamics and Control.* 2(February 1980): 109-32.

Actual Competition, Potential Competition, and Bank Profitability in Rural Markets

by Gary Whalen

Gary Whalen is an economic advisor at the Federal Reserve Bank of Cleveland. The author wishes to acknowledge the helpful comments of Kelly Eakin.

Introduction

The nature of the relationship between the structure of the market in which banks operate—the number and size distribution of actual competitors in a market—and their performance has been examined in a considerable number of empirical studies over the past 20 years.[1] Industrial organization economists have investigated the structure/performance relationship for a wide variety of intra- and interindustry samples of firms.

The typical maintained hypothesis has been that explicit or tacit collusion is more likely in markets with a limited number of large competitors and should result in a statistically significant positive relationship between market concentration and the profitability of firms operating in the market. Definitive support for this hypothesis implies that an activist antitrust policy aimed at limiting merger-related increases in concentration is an appropriate public policy goal.

A positive concentration/profits relationship has been found in some, but far from all, of the empirical studies investigating bank market structure and performance. The mixed results of this body of empirical work have been interpreted in widely different ways.

Some researchers, predisposed to accept the reasonableness of the concentration/collusion hypothesis, have concluded that the weight of the evidence supports this position and have advanced a number of reasons to discount the lack of consistent empirical support for the expected relationship between concentration and bank profitability.[2] One is that the equations estimated in many of these studies have been misspecified, possibly biasing the estimated coefficient on the concentration variable. In particular, several researchers have suggested that market concentration might impact bank management's risk-return preferences or opportunities.[3] Specifically, bank management operating in concentrated markets might trade off potential monopoly profits for lower risk. If this is the case, significant concentration-related differences in profitability might not be evident in studies that fail to explicitly control for risk.

Other researchers have argued that the single-equation estimation techniques typically used in previous empirical work, even those where risk measures have been included as additional

■ 1 For reviews of this work, see Rhoades (1982), Gilbert (1984), and Osborne and Wendel (1983).

■ 2 This is the conclusion of Rhoades (1982).

■ 3 See Heggestad (1977), Rhoades and Rutz (1982), Clark (1986b), and Liang (1987).

explanatory variables, may have biased the results.[4] In their view, profitability and risk are determined simultaneously, so we should rely only on the results of studies where the relationships between these variables and concentration are investigated using simultaneous equation estimation techniques.

Yet another group of researchers argue that the concentration/collusion hypothesis is unreasonable because it embodies a questionable implicit assumption: that technological conditions, regulation, other barriers to entry, or the threat of predation allow colluding firms in concentrated markets to disregard potential competitors.

Concentration-related monopoly power and profits can exist and persist only when there is no threat of entry by potential competitors.[5] Markets in which this type of behavior can occur have been given the label "noncontestable." In theoretical work, researchers have shown that when entry and exit are not precluded, or a market is contestable, then outcomes can approximate those of perfect competition even if the number of actual competitors is quite small or if concentration is high.[6] Consequently, firm profitability should not be expected to vary with concentration.

The possibility that potential competitors may significantly affect the prices charged and profits earned by incumbent firms has been recognized for some time.[7] Until quite recently, however, banks and other financial intermediaries faced numerous regulatory and legislative constraints on geographic location, on permissible products and services they could offer, as well as on the prices they could charge. Thus, few of the geographic and product markets in which banks operated approximated the contestable ideal.

This situation has changed dramatically in the past 10 years. A large number of states have reduced intrastate and, more recently, interstate barriers to geographic expansion by commercial banks and by savings and loan institutions. In addition, the repeal of usury laws and removal of Regulation Q ceilings on deposit rates have left financial intermediaries basically free to compete on a price basis.

Empirical investigations of scale and scope economies in banking suggest that small-scale entry is not precluded by cost conditions.[8] A negligible amount of the costs of branching appears to be sunk. These circumstances suggest that banking markets—at least in states that have liberalized branching to some extent, facilitating entry by out-of-market firms—have become contestable. Alternatively, potential competition may have become an effective disciplinary force, which could explain the absence of a strong positive concentration/profitability relationship in some of the more recent empirical studies.[9]

Researchers who do not subscribe to the concentration/collusion hypothesis have offered an alternative explanation for the significant positive relationship between concentration and profitability reported in some previous studies. They argue that such a finding need not necessarily signal collusion or indicate causation running from concentration to profitability. In their view, labeled the "efficient structure hypothesis" (ESH), superior efficiency, management, or luck could result in increased firm profitability and market share and, ultimately, in higher concentration.[10] If the ESH is correct, then the positive relationship between concentration and profitability detected in empirical work where a market share variable is not included is spurious and simply reflects the correlation between market share and concentration.

At present, then, there continues to be a great deal of uncertainty and disagreement about the relationship between market concentration, potential competition, and bank performance. Very few of the numerous previous studies have incorporated risk, controlled for market share, and investigated possible simultaneity.

More important, virtually no empirical work on the impact of potential competition in banking, or in any other industry for that matter, has

■ **4** This is the conclusion of Clark (1986b) and Liang (1987).

■ **5** See Brozen (1982) and Baumol, Panzar, and Willig (1982).

■ **6** Actually, researchers have differentiated markets according to the degree to which they are contestable. At one extreme are noncontestable markets. At the other extreme are perfectly contestable markets. In essence, perfectly contestable markets are ones in which entry and exit are costless. This, in turn, implies no barriers of any kind to entry and exit. In particular, zero sunk costs are required to enter the market. Markets in which entry and exit can occur but are not costless have been labeled imperfectly contestable. In such markets, potential competition is expected to influence the performance of incumbent firms. For a more detailed discussion of these issues, see Schwartz (1986), pp. 37-48, and Morrison and Winston (1987), pp. 53-60.

■ **7** This possibility was noted in Bain (1949) more than 30 years ago.

■ **8** See Berger, Hanweck, and Humphrey (1986).

■ **9** For example, Evanoff and Fortier (1988) find evidence of a positive concentration/profitability relationship for a subsample of banks drawn from unit banking states but not for the subsample drawn from states where branching is permitted.

■ **10** See Smirlock (1985).

been done to date.[11] A number of circumstances make banking an ideal subject for such research. The partial, gradual elimination of geographic barriers to market entry, cost conditions, and the local nature of banking markets mean that entry can occur if market conditions warrant and that the number of potential bank entrants for each local market can be determined.

This paper attempts to provide more definitive evidence on the relationship between competition and bank profitability. The relationship between bank profitability and both actual and potential competition is examined in a framework that explicitly includes market share and risk variables. Further, the impact of possible simultaneity is also explored.

The sample consists of 159 banks drawn from non-MSA (metropolitan statistical area) counties in Ohio. The focus is on non-MSA counties for several reasons. First, the number of actual bank competitors in a typical non-MSA county is generally small, and concentration is high relative to MSAs in the state. Second, economic and demographic characteristics of rural counties generally make them less attractive for entry than urban counties. Finally, actual and potential competition from out-of-market and nonbank suppliers of financial services is likely to be limited.

Thus, if the concentration/collusion hypothesis is correct and if potential competition is a relatively unimportant determinant of firm performance, supporting empirical evidence is likely to be obtained from this data set. Conversely, absence of support for the concentration/collusion hypothesis and the finding that potential competition impacts bank performance in rural markets is strong evidence that local banking markets, both rural and urban, are contestable.

The time interval examined is from 1979 to 1981. This particular period was chosen because the bank branching law in Ohio was liberalized in January 1979. Before then, de novo branching was limited to a bank's home office county. Under the new law, banks could branch de novo into all counties contiguous to the county in which their head office was located. Thus, the partial removal of geographic restrictions on branching created an identifiable number of potential bank entrants for each county in the state.

The choice of a three-year time period appears somewhat arbitrary. However, a period of this length should be short enough to ensure that ongoing expansion activity by banks does not materially affect the measure of potential competition used in the study. It should also be long enough to allow any performance impacts attributable to potential competition to be detected statistically.

In the following sections, we discuss the model to be estimated, describe the sample and estimation techniques, and present the results. A summary and conclusions follow.

I. Model Specification

Unfortunately, there continues to be no strong consensus about the "best" microeconomic model of the banking firm. As a result, researchers disagree about how the profitability equation to be estimated—whether a single reduced-form equation or a structural equation in a simultaneous system—should be specified. No attempt is made here to resolve the theoretical debate. Our approach is simply to estimate versions used in previous studies, with market share, risk, and potential competition variables explicitly included.

Thus, the profitability equations estimated had the following general form:

$$(1)\quad PROF_i = f(AC_i, PC_i, MS_i, RISK_i, \underline{Z}_i)$$

where

$PROF_i$: a measure of the profitability of bank i

AC_i: a proxy for actual competition in the market in which bank i operates

PC_i: a proxy for potential competition faced by bank i

MS_i: the market share of bank i

$RISK_i$: a measure of the overall risk of bank i

$\underline{Z}_i$: a vector of additional control variables

The profitability measure employed as the dependent variable in this study is rate of return on equity (net income after taxes, excluding securities gains and losses, divided by book equity,

■ 11 The only explicit empirical test to date is Hannan (1979). In many structure/performance studies, the sign and statistical significance of coefficients on branching law dummies in estimated profitability equations are used to draw inferences about the intensity of potential competition. In others, the statistical significance (or lack of significance) of the estimated coefficient on the concentration term is used to obtain insight on this issue. In fact, very few explicit empirical tests of contestability/potential competition have been done for any industry, including the airline industry, which Baumol, et al. cited as an example of one with contestable markets. The study by Morrison and Winston (1987) may be the only one published to date.

both measured at year-end) averaged over the three years from 1979 to 1981. This profitability measure best reflects the efforts of managers interested in shareholder wealth maximization.

The determinants of profitability of primary interest in this study are actual and potential competition. The former is proxied in two alternative ways: by incumbent-firm market concentration and by the number of actual competitors. The latter is proxied only by the number of potential competitors.[12]

The precise form of the relationship between the proxies for actual competition, potential competition, and profitability are unclear and could take a number of different forms.

The consensus view is that actual competition will be more intense and incumbent profitability will be lower, the greater the number of actual competitors or the lower the market concentration. The relationship between these proxies, the likelihood of collusion, and the intensity of competition and ultimately profitability might not be linear, however.[13] For example, the marginal impact of additional actual competitors might not be constant, but could decline as the number of competitors increased. As a result, we also investigate nonlinear relationships between the proxies for actual competition and profitability.

As long as entry into rural banking markets is not precluded, the prices and profits of incumbents should also vary systematically with the number of potential entrants. However, there is some uncertainty about the precise form of the relationship between incumbent profitability and the *number* of potential competitors because the relationship between the number of potential competitors and the intensity of potential competition is unclear.[14] The standard view appears to be that the larger the number of potential entrants, the greater the perceived threat of entry and the lower the incumbent prices and profits.

Some writers, however, have suggested that when more than one potential entrant exists, each potential entrant will recognize that entry by others could occur and could impact its expected profit.[15] Researchers have demonstrated that mutual awareness among potential entrants could cause the relationship between the number of potential entrants and the overall likelihood of entry to be non-monotonic, perhaps even negative. This type of relationship implies that the negative marginal impact of additional potential competitors on incumbent profitability could decline as the number of potential entrants increases. Because of this possibility, a quadratic potential competition specification is also explored.

Several researchers have also suggested that the impact of potential competition could vary with the intensity of actual market competition, and possibly with the two measures of market structure employed here to proxy this force.[16] In particular, a given number of potential competitors could impose a larger impact on incumbent profitability if actual competition in the market were less intense. To investigate this possibility, actual competition/potential competition interaction variables are included in several versions of the performance equations estimated.

Our study uses two summary measures of incumbent market structure: the three-firm deposit concentration ratio and the number of actual competitors. Two variants of each of these measures are employed. One is calculated using data for commercial banks only. The other is calculated using data for both banks and savings and loans, in recognition of the typically considerable thrift share of deposits in counties throughout Ohio and their expanding ability to compete with commercial banks.

The number of holding company organizations legally permitted to branch de novo into each market is the measure of potential competition employed in this analysis. Available data revealed that holding company affiliates were responsible for most of the de novo branching activity in Ohio from 1979 to 1981. We exclude smaller banks that are unlikely to branch de novo in order to produce a more precise measure of potential competition.[17]

■ **12** Since it is not clear that the size distribution of potential competitors influences their performance impact, and since construction of a measure of potential competitor concentration would be extremely tedious, only the number of potential competitors is employed.

■ **13** The possibility of a nonlinear relationship between measures of market structure and performance is noted in Heggestad (1979), pp. 468-69.

■ **14** For a discussion of the expected relationship between concentration, potential competition, and incumbent profitability, see Call and Keeler (1986), p. 224; Schwartz (1986), pp. 47-48; and Morrison and Winston (1987).

■ **15** See Kalish, Hartzog, and Cassidy (1978). Empirical evidence supporting this view appears in Hannan (1981) and Morrison and Winston (1987).

■ **16** Possible interactions between measures of actual and potential competition are discussed in Hannan (1979), pp. 442-43, and in Morrison and Winston (1987), p. 63.

■ **17** Examination of data on branching in Ohio over the 1979 to 1981 period revealed that holding company affiliates established 61 percent of the total number of de novo branches over this interval. Further, they established 64 percent of those opened in contiguous counties. See Whalen (1981).

Following the approach taken with the concentration variable, market share for each bank is defined in two different ways: by its share of commercial bank deposits in the market and by its share of bank and savings and loan deposits in the market. An insignificant coefficient on the incumbent market structure variable, in conjunction with a positive, significant coefficient on the related market share term, is evidence supporting the efficient structure hypothesis.

The risk measure used in this study is the same one used by a number of previous researchers: the standard deviation of return on equity over the period examined (1979 to 1981). There is some disagreement about the nature of the relationship between this variable and profitability. Heggestad (1979) and Clark (1986b) have argued that the relationship should be positive; Liang (1987) has suggested that it should be negative.[18] There is empirical evidence in support of both positions. Because of the uncertainty and because the precise nature of the relationship between these two variables is not the primary focus of this paper, the anticipated sign of the coefficient on the risk measure is left ambiguous.

The other explanatory variables in the estimated profitability equations are elements of the vector, $\underline{Z}$. These are presumably exogenous variables that reflect differences in the characteristics of an individual bank, or economic conditions in its market or its regulatory environment that could influence its profitability.

Three bank characteristic variables are employed: a bank size measure, a dummy variable measure of the number of branches operated, and a dummy variable indicating whether the bank was a subsidiary of a bank holding company. Economic conditions in each bank's local market are represented by two variables: average per capita personal income and per capita personal income growth. Finally, we use a Federal Reserve System membership dummy to control for regulation-related cost differentials.

To determine if the estimated relationship between actual competition, potential competition, and profitability is materially influenced by the neglect of possible simultaneity, the profitability equation is also viewed as a structural equation in a multi-equation simultaneous system. Specifically, a two-equation system similar to that used in Liang (1987) is employed. In this system, bank risk is the other endogenous variable. The main difference between her specification and the one employed here is the addition of the potential competition term.

Liang's structural equation for risk contains five predetermined variables that do not appear in the profitability equation discussed above. These variables are designed to proxy market uncertainty. They are the standard deviation of market per capita personal income, unexplained market deposit supply, unexplained variation in bank i's loan demand, unexplained variation in bank i's deposit supply, and the covariance of bank i's unexplained loan demand and deposit supply. The precise definition of each of these variables and the reduced-form equations for this model are detailed in the appendix.

II. Sample and Methodology

Our sample consists of the 159 single-market banks headquartered in non-MSA Ohio counties at the end of 1981. Single-market banks are those with all offices located within their home office county. This criterion allows their performance to be related to the characteristics of their particular local markets. The presumption is that non-MSA counties approximate local rural banking markets.

The profitability equations are estimated using two different statistical techniques. Ordinary least squares regression (OLS) is used to estimate versions in which risk is viewed as exogenous. Two-stage least squares (2SLS) is the technique used to estimate the profitability equation when it is viewed as part of a simultaneous system.

III. Results

Regression results are presented in tables 1 and 2. Only the equations containing measures of actual market structure and market share calculated using commercial bank data are included in the tables. The results were essentially the same when savings and loans were considered in the calculation of these variables and therefore are not reported.

Table 1 contains versions of the profitability equation estimated using OLS; table 2 contains abbreviated results obtained by estimating versions of the equations in table 1 viewed as part of a two-equation simultaneous model. The estimation technique is 2SLS. Only the coefficients and t-statistics for the actual competition, potential competition, market share, and risk variables are reported. In general, the overall explanatory

■ **18** In Liang's model, greater profit variability implies greater expected costs and associated penalties to the bank, resulting in a negative relationship between profit variability and expected profit margins.

TABLE 1

OLS Versions of Profitability Equations
Dependent Variable: *AROE*

Variables	(1) Coefficient	(2) Coefficient	(3) Coefficient	(4) Coefficient
CBO	-0.003253	-0.012121		
	(-0.17)	(-0.62)		
NCBO			0.018071	-0.520431
			(0.12)	(-1.80)
MSBO	0.036970	0.043162	0.036723	0.035676
	(1.58)	(1.85)	(1.48)	(1.45)
PCPIGR	0.096151	0.100644	0.094267	0.090787
	(1.26)	(1.33)	(1.20)	(1.17)
PCPI	0.000213	0.000204	0.000210	0.000175
	(0.80)	(0.77)	(0.78)	(0.66)
OD	-0.495045	-0.634474	-0.496847	-0.467602
	(-0.76)	(-0.98)	(-0.76)	(-0.72)
FRM	-0.003253	-0.149174	-0.031221	-0.095712
	(-0.17)	(-0.28)	(-0.06)	(-0.18)
MBHC	2.183394	2.113173	2.186579	2.271347
	(3.14)	(3.06)	(3.10)	(3.26)
SIZE	-0.741219	-0.791260	-0.734965	-0.834644
	(-1.49)	(-1.61)	(-1.46)	(-1.68)
SDROE	-0.757202	-0.737778	-0.757808	-0.750641
	(-8.07)	(-7.91)	(-8.06)	(-8.08)
HCPE	-0.158573	-1.219721	-0.156887	-0.806247
	(-1.29)	(-2.34)	(-1.28)	(-2.46)
HCPESQ		0.114109		
		(2.10)		
HCNCBO				0.112784
				(2.16)
INT	14.110421	16.894396	13.787394	17.385094
	(4.31)	(4.83)	(4.97)	(5.42)
F	7.46	7.34	7.41	7.33
RSQ	0.34	0.35	0.34	0.36

NOTE: T-statistics are in parentheses.
SOURCE: Author.

power of the estimated equations is good, given the size and cross-sectional nature of the sample.

The coefficients on the actual and potential competition and market share variables are of primary interest. The signs and statistical significance of the other variables in the estimated equations are of secondary importance here and will not be discussed.

The coefficient on the concentration variable is never even marginally significant in any version of the equation estimated.[19] The results were invariant to specification and estimation techniques. Including savings and loans in the calculation of this variable and excluding the market share term did not alter this finding.

When the number of actual competitors is used as the actual competition proxy, the results obtained do vary with the specification employed. The coefficient on the number of actual competitors term is insignificant when a linear specification is employed and when an actual competition/potential competition interaction term is not included in the estimated equation. However, when an interaction term is included, the coefficient on the number of actual competitors variable becomes negative and significant. This result holds when savings and loans are included in this measure and when a simultaneous-equations estimation technique is employed. The coefficients are not significant when a quadratic version is examined.

The estimated coefficient on the number of potential competitors variable is negative, but only marginally significant (that is, 10 percent level, one-tail test) when a linear specification is employed and when an actual competition/potential competition interaction term is not included. However, when this variable is used in an estimated equation in conjunction with the number of actual competitors and an interaction term, the coefficient is negative and significant.

In these equations, the actual competition/potential competition interaction term, constructed by multiplying the number of actual and potential competitors, exhibits a positive significant coefficient. This finding supports the view that the negative marginal impact of additional actual competitors declines as the number of potential competitors increases. Similarly, the larger the number of actual competitors in a market, the smaller the negative marginal impact of additional potential competitors.

When a quadratic potential competition specification is employed, the estimated coefficients on the number of potential competitors term and the square of this variable are both significant. The pattern of signs (negative and positive, respectively) could reflect mutual awareness among potential entrants. This result suggests that the marginal impact of additional potential competitors is initially negative.

■ **19** A Herfindahl-Hirschmann Index of market concentration was also employed in place of the three-firm concentration ratio. The change in the definition of the concentration ratio did not materially impact the results.

However, the size of the negative impact declines as the number of potential competitors increases and finally turns positive. The magnitudes of the coefficients imply that incumbent firm profitability is constrained in markets with five or fewer potential entrants. This finding supports the notion of a nonlinear relationship between the number of potential entrants and the overall probability of entry.

Changing the definition of the market structure and market share variables to include savings and loans did not alter either the size or the statistical significance of the coefficients on the potential competition variables in any of the specifications examined. Further, a comparison of each equation in table 1 with its counterpart in table 2 also demonstrates that the sign and statistical significance of the coefficients on the variables of interest in the estimated equations are not sensitive to the estimation technique used.[20] This was true for the other exogenous control variables as well.

TABLE 2

Summary Results
2SLS Versions of Profitability Equations
Dependent Variable: *AROE*

Variables	(1) Coefficient	(2) Coefficient	(3) Coefficient	(4) Coefficient
CBO	-0.001529 (-0.08)	-0.010870 (-0.53)		
NCBO			0.007936 (0.05)	-0.520134 (-1.79)
MSBO	0.036002 (1.52)	0.042391 (1.79)	0.035831 (1.43)	0.035265 (1.43)
$\widehat{SDROE}$	-0.857773 (-2.95)	-0.810309 (-2.78)	-0.870202 (-3.00)	-0.803704 (-2.80)
HCPE	-0.159169 (-1.28)	-1.186875 (-2.17)	-0.158830 (-1.30)	-0.801043 (-2.46)
HCPESQ		0.110422 (1.93)		
HCNCBO				0.111721 (2.13)
F	1.80	2.33	1.81	2.11
RSQ	0.11	0.15	0.11	0.14

NOTE: T-statistics are in parentheses.
SOURCE: Author.

In general, the coefficient on the market share variable is positive and at least marginally significant (at the 10 percent level, one-tail test) in every variant of the profitability equation estimated. As with the concentration measure, somewhat stronger results are obtained when savings and loan deposits are considered in the construction of this variable.

IV. Summary and Conclusions

The results support the notion that non-MSA banking markets are contestable. That is, we found bank performance to be systematically related to proxies designed to measure the intensity of actual and potential competition. The threat of entry by potential competitors does appear to limit incumbent firm profitability, although the threat of entry and the number of potential competitors may not be monotonically related. Incorporating risk into the analysis and considering possible simultaneity between risk and profitability did not materially alter the results.

Both proxies for actual competition were not found to be consistently related to bank performance, however. The concentration measure was not found to be significantly related to the profitability of banks operating in rural markets in Ohio in any specification investigated. Only the number of competitors proxy was found to be significantly related to bank profitability in the expected way.

The finding that potential competition has a significant impact on incumbent performance is somewhat surprising for several reasons. First, potential competition is generally expected to be a weak force in rural banking markets. Second, researchers have argued that potential entrants may not significantly impact incumbent prices and profits in periods immediately after a change in regulations that affects entry conditions. The interval analyzed was just such a period. In addition, the potential entrant variable used in this study does not include potential nonbank competitors, particularly savings and loans. Thus, the variable is obviously not a perfect proxy for the threat of entry in the markets examined.

Further research on the impact of potential competition in banking markets appears warranted to determine if the observed relationships

■ **20** In addition, to further examine the sensitivity of the results to changes in specification, versions of the profitability equation similar to the one appearing in the four-equation model developed in Clark (1986b) were also estimated. The only change in Clark's specification was the addition of the potential competition measures used in this study. Again, this change in specification did not materially alter the results reported above.

are evident for other samples of banks and in other time periods. However, the results of this study suggest that it is unclear whether the consolidation taking place in banking in recent years has substantially lessened competition, given the simultaneous reductions in barriers to market entry that have occurred.

For bank regulatory agencies, the results also imply that the competitive impacts of bank mergers cannot be reliably determined solely from a mechanical analysis of changes in actual market structure. Entry conditions and the existence of potential competition should also be considered and used to temper conclusions drawn from an analysis of merger-related changes in concentration or in the number of actual competitors.

APPENDIX

Variable Definitions

AROE: Bank *i* 's annual after-tax return on equity, averaged over the 1979-1981 period.

CBO: Three-firm market concentration ratio, banks only, June 1980.

NCBO: Number of banks operating in the market of bank *i,* June 1980.

HCPE: Number of holding company organizations legally permitted to branch de novo into the market.

HCPESQ: The square of *HCPE.*

HC——: Interaction term. *HCPE* times various alternative measures of market structure.

MSBO: Bank *i* 's deposit market share, banks only, June 1980.

SDROE: Bank *i* 's standard deviation of annual after-tax return on equity over the 1979-1981 period.

SIZE: Log of total assets of bank *i.*

OD: Dummy variable equal to one if bank *i* has at least one branch, otherwise equal to zero.

FRM: Dummy variable equal to one if bank *i* was a member of the Federal Reserve System, otherwise equal to zero.

MBHC: Dummy variable equal to one if bank *i* is a holding company subsidiary, otherwise equal to zero.

PCPI: Per capita personal income in the market averaged over the 1979-1981 interval.

PCPIGR: Per capita personal income growth in the market over the 1979-1981 interval.

SDPCPI: The standard deviation of market per capita personal income over the 1979-1981 interval.

MDU: Market deposit uncertainty variable equal to proportion of unexplained variation in market deposits derived from the regression of market deposits on market income over the 1979-1981 interval.

LRISK: Loan uncertainty variable for bank *i* equal to proportion of unexplained variation in total loans derived from the regression of total loans on market income over the 1979-1981 interval.

DRISK: Deposit uncertainty variable for bank *i* equal to proportion of unexplained variation in total transactions deposits derived from the regression of total transactions deposits on market income over the 1979-1981 interval.

COVLD: Covariance of unexplained loans and deposits for bank *i* over the 1979-1981 period.

$\widehat{SDROE}$: Predicted value for *SDROE* derived from the following first-stage regression with the relevant actual and potential competition variable(s) added:

$$\widehat{SDROE} = f\,(MSBO, SIZE, OD, FRM, MBHC, PCPI, PCPIGR, SDPCPI, MDU, LRISK, DRISK, COVLD).$$

References

Bain, Joe S. "A Note On Pricing in Monopoly and Oligopoly." *American Economic Review.* 39(March 1949): 448-64.

Baumol, William J., John C. Panzar, and Robert D. Willig. *Contestable Markets and the Theory of Industry Structure.* New York: Harcourt, Brace and Jovanovich, 1982.

Berger, Allen N., and Timothy Hannan. "The Price-Concentration Relationship in Banking." In *Merging Commercial and Investment Banking: Proceedings from a Conference on Bank Structure and Competition.* Federal Reserve Bank of Chicago (May 1987): 538-9.

Berger, Allen N., Gerald A. Hanweck, and David B. Humphrey. "Competitive Viability in Banking: Scale, Scope, and Product Mix Economies." Board of Governors of the Federal Reserve System. *Research Papers in Banking and Financial Economics,* February 1986.

Boyle, Stanley E., Richard K. Ford, and Charles G. Martin. "The Application of Industrial Organization Theory to Commercial Banking: A Review and Analysis." University of Arkansas-Little Rock, Department of Economics and Finance. *Working Paper,* 1987.

Brozen, Yale. *Concentration, Mergers, and Public Policy.* New York: Macmillan Publishing Company, Inc., 1982.

Burke, J., and S. Rhoades. "Profits and 'Contestability' in Highly Concentrated Banking Markets." *Review of Industrial Organization.* (Fall 1987).

Call, C., and T. Keeler. "Airline Deregulation, Fares, and Market Behavior: Some Empirical Evidence." In A. Daughety, ed., *Analytical Studies in Transport Economics.* Cambridge, Mass.: Cambridge University Press, 1986.

Clark, Jeffrey A. "Single-Equation, Multiple-Regression Methodology: Is It an Appropriate Methodology for the Estimation of the Structure-Performance Relationship in Banking?" *Journal of Monetary Economics.* 18, no. 3(November 1986a): 295-312.

________. "Market Structure, Risk, and Profitability: The Quiet Life Hypothesis Revisited." *Quarterly Review of Economics and Business.* 26, no. 1(Spring 1986b): 45-56.

Evanoff, Douglas D., and Diana L. Fortier. "Reevaluation of the Structure-Conduct-Performance Paradigm in Banking." *Journal of Financial Services Research.* 1, no. 3(June 1988): 277-94.

Gilbert, R. "Bank Market Structure and Competition." *Journal of Money, Credit, and Banking.* 16, no. 4(November 1984).

Graddy, Duane B., and Reuben Kyle III. "The Simultaneity of Bank Decision-Making, Market Structure, and Bank Performance." *Journal of Finance.* 34, no. 1(March 1979): 1-18.

Graham, David R., Daniel P. Kaplan, and David S. Sibley. "Efficiency and Competition in the Airline Industry." *Bell Journal of Economics.* 14, no. 1(Spring 1983): 118-38.

Hannan, Timothy H. "Mutual Awareness Among Potential Entrants: An Empirical Examination." *Southern Economic Journal.* 47, no. 3(January 1981): 805-8.

________. "Limit Pricing and the Banking Industry." *Journal of Money, Credit, and Banking.* 11, no. 4(November 1979): 438-46.

Heggestad, Arnold A. "Market Structure, Competition, and Performance in Financial Industries: A Survey of Banking Studies." In Franklin R. Edwards, ed., *Issues in Financial Regulation.* New York: McGraw Hill Book Company, 1979.

________. "Market Structure, Risk, and Profitability in Commercial Banking." *Journal of Finance.* 32, no. 4(September 1977): 1207-16.

Kalish, Lionel, Jerry Hartzog, and Henry Cassidy. "The Threat of Entry With Mutually Aware Potential Entrants: Comment." *Journal of Political Economy.* 86, no. 1(February 1978): 147-50.

Liang, Nellie. "Bank Profitability and Risk." Unpublished paper, Board of Governors of the Federal Reserve System, November 1987.

Morrison, Steven A., and Clifford Winston. "Empirical Implications and Tests of the Contestability Hypothesis." *Journal of Law and Economics.* 30(April 1987): 53-66.

Osborne, Dale K., and Jeanne Wendel. "Research on Structure, Conduct, and Performance in Banking, 1964-1979." Oklahoma State University, College of Business Administration. *Working Paper* 83-8, July 1983.

Rhoades, Stephen A. "Structure-Performance Studies in Banking: An Updated Summary and Evaluation." Board of Governors of the Federal Reserve System, Staff Study No. 119, August 1982.

________ , **and Roger D. Rutz.** "Market Power and Firm Risk: A Test of the 'Quiet Life' Hypothesis." *Journal of Monetary Economics.* 9(January 1982): 73-86.

Schwartz, Marius. "The Nature and Scope of Contestability Theory." *Oxford Economic Papers.* 38, Supplement (November 1986): 37-57.

Smirlock, Michael. "Evidence on the (Non)Relationship Between Concentration and Profitability in Banking." *Journal of Money, Credit, and Banking.* 17, no. 1(February 1985).

Whalen, Gary. "Concentration and Profitability in Non-MSA Banking Markets." Federal Reserve Bank of Cleveland. *Economic Review* (Quarter 1, 1987): 2-14.

________ . "Bank Expansion in Ohio." Federal Reserve Bank of Cleveland. *Economic Commentary,* April 6, 1981.

Getting the Noise Out: Filtering Early GNP Estimates

by John Scadding

John Scadding was a visiting scholar at the Federal Reserve Bank of Cleveland when he wrote this paper. Currently, he is an economist with the California Public Utilities Commission. The author would like to thank Stephen McNees for helpful comments.

Introduction

Real, or inflation-adjusted, gross national product (GNP) is the most inclusive measure of the nation's economic activity. As such, it is probably the most closely monitored economic barometer for the information it contains about the economic well-being of the economy and about the economy's prospects. It is the central focus of most macroeconomic models and their forecasts, and it plays a decisive role in shaping monetary and fiscal policy decisions.

Given the critical role that GNP plays, it is not surprising that the accuracy of GNP estimates is crucial if informed decisions are to be made by both private agents and government policymakers. There is a trade-off, however, between the estimates' accuracy and their *timeliness.* Delays in reporting and revising data as more inclusive information becomes available means later estimates will typically be more accurate than earlier ones; but waiting longer entails forgoing the opportunity to take action sooner, when that may be a critical factor.

In the United States, the first offical estimate for a particular quarter's GNP is released by the U. S. Department of Commerce approximately three weeks *after* that quarter has ended. Much of the data needed to construct GNP are still not available at that point, even though the quarter has ended. The missing data therefore must be estimated by the U.S. Department of Commerce's Bureau of Economic Analysis (BEA), which is responsible for compiling the official estimate of GNP. This first estimate is followed in relatively rapid succession by two additional estimates, one and two months after the initial number is released. Thereafter, the delays in revisions become much longer. Estimates are usually subject to three further annual revisions. After that, an estimate is usually subject to further so-called benchmark revisions every five years as data from the Bureau of Census' quinquennial economic census are incorporated. At each stage, source data are incorporated that had not been available previously, and revisions to previous data are incorporated as well.[1]

It is clear from this description that there is never a final estimate of GNP that could be equated with the "truth." Nevertheless, the three early preliminary or *provisional* estimates are obviously distinct from the later ones in terms of their timeliness. Although based on incomplete and preliminary information, the provisional estimates have the advantage that they are available

■ 1 Carson (1987) provides a comprehensive overview of the source data and estimation methods for constructing the different GNP estimates. See also Young (1987).

much sooner than the later, more comprehensive, and presumably more accurate numbers. It is relevant, therefore, to examine their accuracy in predicting the later numbers. As Allan Young, director of the Bureau of Economic Analysis, noted in a recent comprehensive survey of the properties of GNP estimates: "Much of the concern with the reliability of GNP comes down to whether the early . . . quarterly estimates . . . provide a useful indicator of the estimates . . . When complete and final source data are available." (Young [1987], p. 18)

TABLE 1

Final Revisions to Real GNP Growth and Components, 1974:IIQ - 1984:IQ

	Final Revisions		Estimated Observation Error		Residual Forecast Error	
	Mean	Variance	Mean	Variance	Mean	Variance
Final minus 15-day	0.630	4.087	−0.630	0.764	0.00	3.323
Final minus 45-day	0.413	2.876	−0.413	0.694	0.00	2.182
Final minus 75-day	0.205	2.742	−0.205	0.890	0.00	1.852

SOURCE: Author.

One important strand of the literature examining this question has concluded that the early numbers can be viewed as *rational* forecasts of the actual numbers. The term *rational* is used in the sense that the differences between a final GNP growth number and its corresponding preliminary estimates are uncorrelated with the preliminary numbers themselves (Mankiw and Shapiro [1986]; Walsh [1985]). On the face of it, this is a surprising result. It denies the intuitively appealing, and perhaps prevalent, view that if a preliminary estimate showed large positive growth for real GNP in a quarter, for example, it would be more likely than not that later estimates would be revised down—in other words, that the final GNP number would be smaller than its preliminary estimate. And, similarly, a large (in absolute value) negative preliminary estimate would be revised upward subsequently.

In a preliminary analysis reexamining this question, Scadding (1987) concluded that the statistical test used in the analyses mentioned above could not discriminate very well between the rational forecasts hypothesis and the alternative view that subsequent revisions to the GNP numbers would be correlated with the preliminary estimates. This alternative view implies that the early GNP numbers are estimates of the final number, but estimates that are contaminated with error.

If this alternative view is correct, then it is possible in principle to make estimates of the error in the preliminary numbers and to adjust the latter to remove the error—in other words, to filter out the "noise." This paper investigates one method of doing this. The results suggest there is scope for adjusting the provisional GNP growth rate numbers to make them better predictors of what the final numbers will turn out to be.

I. The Data

Table 1 has estimates of the *final revisions* for real GNP growth—that is, the difference between the final estimate and the three provisional estimates. There are three final revisions, corresponding to the difference between the final numbers and each of the three provisional numbers. For the sample period used in this paper (1974-1984), the early estimates came out 15 days, 45 days, and 75 days after the quarter ended, and the usual nomenclature is to refer to them as the 15-day estimate, and so on. Correspondingly, there is the 15-day final revision, which is the difference between the final number and the 15-day estimate, and so on. I follow the usual practice and define the "final" number as the currently available final number as of the quarter in question. Thus, final estimates in the earlier part of the sample will have been through more revisions than those later in the sample.[2]

For the 15-day estimate of GNP, many of the source data are not complete and are subject to revision. The data available for this estimate are monthly data, like retail sales, manufacturers' shipments of machinery and equipment, and merchandise trade figures. Some of these data, like retail sales, are based on surveys, and typically are revised substantially. In addition, some of the monthly source data are not available for all three months of the quarter. For example, only one to two months of data are available for estimating consumer spending on services, which is about one-half of total consumer spending. And there are no monthly data at all for

■ 2 The data are from a study prepared by the Bureau of Economic Analysis and are the data used by Mankiw and Shapiro (1986), Mork (1987), and Walsh (1985). The data were adjusted to abstract from the effects of definitional changes and the change in the base year for calculating constant-dollar GNP. See Young (1987), p. 25. I am indebted to Professor Mork for providing me with a copy of these data.

about 40 percent of spending on services. This component, therefore, is estimated by the Department of Commerce, either by extrapolating by related series or by judgmental projection.

The succeeding 45- and 75-day estimates incorporate new monthly data unavailable for the 15-day estimate, and as well incorporate revisions to the monthly data that were included in the 15-day number. As well, these two estimates include new information available only on a quarterly basis—domestic corporate profits, balance of payments figures, and data on financial assets from the Federal Reserve Board's flow of funds accounts. The latter two sources are incorporated in the 75-day estimate only (Carson [1987], p. 107).

As table 1 shows, the final revisions are not trivial. On average for the sample they are positive, suggesting a systematic tendency for the preliminary numbers to understate the final estimates, a phenomenon that has been noted elsewhere (Mork [1987]). The deviations implied by the sample variance estimates reported in table 1 are large when measured against the the mean growth of real GNP for the period, which was 2.9 percent. Thus, plus or minus one standard error about a preliminary estimate equal to this trend growth translates into an economy that, with equal probability, could be enjoying near boom-like conditions or behaving as if it was close to recession.

II. The Nature of the Provisional GNP Estimates

As discussed briefly in the introduction, one possible way of thinking of the early GNP growth numbers is as *forecasts* of what the final estimate will turn out to be. Thus, suppose X_t^* is the final estimate of GNP growth for quarter *t;* that estimate of course will not be made until some time after quarter *t.* In the meantime, however, a provisional estimate (in fact three), call it X_t, will be available soon after quarter *t* has ended. This provisional estimate X_t can be thought of as a forecast of what X_t^* will be. From that perspective, it is natural to ask whether X_t is a *good* forecast in the sense that, at a minimum, it is unbiased and is uncorrelated with the forecast error, which is equal to the final revision, $X_t^* - X_t$. If this description fits X_t, then

$$(1) \quad X_t^* = X_t + z_t,$$

where z_t is a zero-mean, serially uncorrelated forecast error (white noise) that is uncorrelated with X_t.

Walsh (1985) defines these to be the properties of a rational forecast. The competing characterization of X_t is that it is an early observation or "reading" of what X_t^* will be, but an observation measured with error. Thus,

$$(2) \quad X_t = X_t^* + u_t$$

where u_t is also white noise, and uncorrelated with X_t^* in this case. Note that this characterization implies that the final revision *is* correlated with the provisional estimate; in other words:

$$(3) \quad \begin{aligned} E(X_t^* - X_t)\, X_t &= -Eu_t^2 \\ &= -\sigma_u^2, \end{aligned}$$

where σ_u^2 is the variance of the observation error *u.*

The evidence on which characterization better describes the nature of the provisional estimates is decidedly mixed. Mankiw and Shapiro (1986) adduce evidence in favor of the position that preliminary numbers are rational forecasts, on the criteria just described. However, I have argued in a technical companion piece to this paper (Scadding [1987]) that their test is likely to have little power. They themselves raise this possibility because of the apparent contradiction of their conclusion with evidence elsewhere that two important data sources for the GNP estimates—retail sales and inventories—have significant measurement errors in them (Howrey [1984] and Conrad and Corrado [1979]).

Walsh, using a slightly different sample from Mankiw and Shapiro, finds corroborating evidence for their result, but this conclusion is compromised by his additional finding that the provisional estimates are inefficient forecasts. In addition, Mork, using different estimation techniques from the other studies, found evidence that the provisional estimates were biased downwards, and that the final revisions were correlated with previous-quarter GNP growth and a forecast of GNP growth from a publicly available survey of private forecasters.

III. Filtering the Early Data

I have argued elsewhere (Scadding [1987]) that Walsh's evidence of inefficient forecasting is equally compatible with the view that provisional GNP numbers are observations rather than forecasts, with the observation errors in the three provisional numbers being sequentially correlated. Howrey (1984) found this to be a useful characterization of the inventory investment component of GNP. In my earlier paper, I devised a test for discriminating between an inefficient forecasts model and a serially correlated measurement error model based on restrictions on the variance-covariance matrix of the final revisions. The results of that test suggest that the provisional estimates of real GNP growth contain measurement error.

The purpose of this paper is to estimate the amount of observation (measurement) error in the provisional GNP growth numbers and subtract that error to obtain modified, or filtered, provisional GNP estimates that have the properties of a rational forecast. Let $\hat{X}_t^*$ be the filtered estimate; then the estimated measurement and forecast errors are defined by

(4a) $\hat{u}_t = X_t - \hat{X}_t^*$ and

(4b) $\hat{z}_t = X_t^* - \hat{X}_t^*$.

The definitions (4a and 4b) implicitly define the decomposition of a final revision, $X_t^* - X_t$, into its measurement and forecast error component:

(5) $X_t^* - X_t = \hat{z}_t - \hat{u}_t$.

Nonrecursive Kalman filtering, described below, is used to specify equations for estimating $\hat{X}_t^*$. Least-squares estimation of these equations yields an $\hat{X}_t^*$ series with the desired forecasting properties:

(6a) $E(X_t^* - \hat{X}_t^*) = 0$ and

(6b) $E(X_t^* - \hat{X}_t^*)\hat{X}_t^* = 0$.

As well, the estimated measurement and forecast errors are orthogonal to each other:

(6c) $E(\hat{u}_t\, \hat{z}_t) = 0$.

Summary statistics for the final revisions and the estimated measurement and forecast errors are shown in table 1. Clearly, the filtering improves the forecasting precision of provisional numbers. The sample variance of the forecasting error after filtering is on the order of 25 to 30 percent lower than the variance of the unfiltered final revision. Nevertheless, the residual forecast variance is still quite large.

The improvement in forecasting precision would appear to be based on two factors. First, the filtered estimates are derived by combining the provisional estimates with a simple time-series forecast of GNP growth. Mork has noted that the prior quarter's GNP has information about the size of the final revision in the current quarter. The time-series forecast presumably is picking up this information. In addition, filtering improves the precision of forecasting by exploiting the fact that part of the final revision is measurement error and therefore can be forecast from the provisional estimates.

Note the uniformly negative means of the estimated observation errors, indicating a systematic tendency of the provisional GNP estimates to underpredict the final numbers. This tendency has been noted by Mork, who ascribes it to concern by the Department of Commerce that the provisional estimates not be seen as being too optimistic and therefore serving some political agenda.

The presence of serially correlated measurement errors makes it relatively easy to predict *interim* revisions—in other words, from the 15-day to the 45-day, and so on—compared to final revisions. As we shall see, the standard errors of the regression predicting the provisional estimates are about 50 percent lower than the standard errors of the equations predicting the final GNP estimates. Thus, the methodology outlined here provides forecasters with a relatively accurate way of forecasting subsequent preliminary estimates. More generally, this result suggests that the provisional estimates are more like each other than they are like the later estimates, a point that has been made by McNees (1986).

Many economists presumably would be offended by the notion that any attention should be paid to forecasting the provisional estimates themselves when what obviously matters is getting a good estimate of the final or "true" number. However, that is "obvious" only to the extent the Federal Reserve or private agents, in reacting to new provisional estimates, discount the measurement error in them, an assumption that is not obvious on its face at least. It is customary to test market forecasts of GNP by their ability to predict final GNP; it would be interesting to inquire whether they do a better job of forecasting provisional GNP estimates.

A final observation suggested by this paper's result is that the frequent practice by forecasters of discarding their GNP growth forecast for a quarter when the first provisional estimate for that quarter becomes available probably is not efficient. The filtering technique used in this paper combines the provisional estimates of GNP growth with a forecast from a simple time-series model. The results suggest that the forecast still has information about final GNP growth even after the preliminary estimates become available. As McNees has noted: "...the distinction between forecasts and 'actual' data is often exaggerated. Both are estimates based on partial, incomplete information." (McNees [1986], p. 3)

IV. The Filtering Framework

The general idea of filtering data is easily sketched out. Suppose the variable we are interested in, X^* (which in our case is the final estimate of GNP growth), evolves over time according to the law of motion

$$(7) \quad X_t^* = \phi X_{t-1}^* + w_t,$$

where ϕ is a fixed parameter and w_t is a random, serially uncorrelated term with zero mean and constant variance (white noise).

We cannot observe X^* directly but have measurements of it, X, that involve error (here X would be a provisional estimate of GNP growth).

$$(8) \quad X_t = hX_t^* + u_t,$$

where h is a fixed parameter and u is also white noise.

The Kalman filter optimally weights the forecast of X^* from equation (7) with the observation to form the best linear unbiased predictor of X^*, called the filtered value:

$$(9) \quad \hat{X}_t^* = \overline{X}_t + K_t(X_t - h\overline{X}_t),$$

where $\overline{X}$ is the forecast and $\hat{X}^*$ is the filtered value. The weighting coefficient, K, is called the *Kalman gain,* and is a function of the variances of w, u, and of h. The filtered value is used to update the forecast. Using (7), this new forecast is combined with the next observation to calculate the next filtered value.

$$(10a) \quad \overline{X}_{t+1} = \phi \hat{X}_t^*$$

$$(10b) \quad \hat{X}_{t+1}^* = \overline{X}_{t+1} + K_{t+1}(X_{t+1} - h\overline{X}_{t+1}).$$

Two modifications are necessary to apply this algorithm to the program at hand. First, the three provisional estimates of GNP growth are repeated observations on the same final estimate. Thus, *within* the quarter, the law of motion is

$$(11) \quad X_{t/n}^* = X_t^*,$$

for $n = 1, 2, 3$, where $n = 1$ refers to the 15-day estimate, $n = 2$ to the 45-day estimate, and $n = 3$ to the 75-day estimate. Similarly,

$$(12a) \quad X_{t/1} = X_t^* + u_{1t},$$

$$(12b) \quad X_{t/2} = X_t^* + u_{2t}, \text{ and}$$

$$(12c) \quad X_{t/3} = X_t^* + u_{3t},$$

where the $X_{t/n}$'s are the provisional estimates and the $u_{n/t}$'s are the corresponding measurement errors. Thus, within the quarter, the ϕ in equation (7) is unity, as is h in (8), while the intraquarter w's are uniformly zero.

The other modification follows from the fact that preliminary estimation suggests that the u's (12) are sequentially correlated. This serial correlation structure is shown in table 2.

The filtering framework is easily adapted to this circumstance by expressing the observation variables in quasi-difference form, $X_{t/3} - a_{12}X_{t/2}$, $X_{t/2} - a_{23}X_{t/1}$, where the a's are the respective serial correlation coefficients of the errors from table 2 (see Bryson and Ho [1969], pp. 400-405). The modified set of filtering equations becomes

$$(13a) \quad \hat{X}_{t/1}^* = \overline{X}_{t/0} + K_1(X_{t/1} - \overline{X}_{t/0}),$$

$$(13b) \quad \hat{X}_{t/2}^* = \hat{X}_{t/1}^* + K_2[(X_{t/2} - a_{23}X_{t/1} - (1-a_{23})\hat{X}_{t/1}^*], \text{ and}$$

$$(13c) \quad \hat{X}_{t/3}^* = \hat{X}_{t/2}^* + K_3[(X_{t/3} - a_{12}X_{t/2}) - (1-a_{12})\hat{X}_{t/2}^*].$$

The initial forecast $\overline{X}_{t/0}$ is taken from a simple time-series model for real GNP growth, X^*. Given the forecast and estimates of the K's and a's, one could then calculate the filtered estimates directly. The approach taken here, however, is to estimate the K's using ordinary least squares to produce a set of estimated measurement errors and residual forecast errors that are uncorrelated.[3] Thus, the estimation equations corresponding to (13) are

$$(14a) \quad X_t^* = \overline{X}_{t/0} + K_1(X_{t/1} - \overline{X}_{t/0}) + z_{1t},$$

$$(14b) \quad X_{t/2} = a_{23} X_{t/1} + (1-a_{23})\hat{X}_{t/1}^* + v_{2t},$$

$$(14c) \quad X_t^* = \hat{X}_{t/1}^* + K_2 v_{2t} + z_{2t},$$

$$(14d) \quad X_{t/3} = a_{12} X_{t/2} + (1-a_{12})\hat{X}_{t/2}^* + v_{3t},$$

and

■ 3 Conrad and Conrado (1979) and Howrey (1984) have used the Kalman framework for analyzing retail sales and inventory investment data, respectively.

$$(14e)\ X_t^* = \hat{X}_{t/2}^* + K_3 v_{3t} + z_{3t}.$$

To complete (14) we append the set of definitions of the filtered estimates of GNP growth:

$$(14a')\ \hat{X}_{t/1}^* = \overline{X}_{t/0} + K_1(X_{t/1} - \overline{X}_{t/0}),$$

$$(14b')\ \hat{X}_{t/2}^* = \hat{X}_{t/1}^* + K_2 v_{2t}, \text{ and}$$

$$(14c')\ \hat{X}_{t/3}^* = \hat{X}_{t/2}^* + K_3 v_{3t}.$$

TABLE 2

Correlation Structure of Final Revisions

	Standard Error of Estimate
$X_t^* - X_{t/3} = \underset{(2.79)}{0.230} + \underset{(19.61)}{0.932}\,(X_t^* - X_{t/2}) + v_{3t}$	0.503
$X_t^* - X_{t/2} = \underset{(0.76)}{0.077} + \underset{(16.36)}{0.784}\,(X_t^* - X_{t/1}) + v_{2t}$	0.503
$X_t^* - X_{t/1} = \underset{(-2.18)}{-0.698} + v_{1t}$	2.02

SOURCE: Author.

The estimation of (14) proceeds sequentially. First (14a) is estimated, by regressing final GNP growth on the time-series forecast, $X_{t/0}$, and the 15-day provisional estimate, $X_{t/1}$. The residual, z_1, is the forecast error for the filtered 15-day estimate. The first filtered estimate of final GNP growth, $\hat{X}_{t/1}^*$, is calculated using (14a′). The corresponding measurement error in the 15-day provisional GNP growth rate is estimated as

$$(15)\quad \hat{u}_{1t} = X_{t/1} - \hat{X}_{t/1}^* = (1-K_1)\,(X_{t/1} - \overline{X}_{t/0}),$$

which by construction is uncorrelated with the forecast error.

The next step is to calculate the innovation in the measurement error in the 45-day provisional GNP number. The correlation structure between the measurement errors in the 15-day and 45-day number is

$$(16)\quad X_{t/2} - X_t^* = a_{23}\,(X_{t/1} - X_t^*) + v_{2t},$$

where v_{2t} is the innovation in the measurement error. Rearranging (16) and substituting $\hat{X}_{t/1}^*$ for X_t^* yields (14b), which is then estimated by regressing the 45-day provisional GNP number on the 15-day estimate and the first filtered estimate of GNP growth.

The innovation in the 45-day number then is used to update the filtered estimate of final GNP growth by regressing the final GNP number on the first filtered estimate and the measurement innovation in the 45-day number (equation 14c). The residual z_{2t} provides an estimate of the forecast error in the 45-day number. The same sequence of estimations is performed to calculate the new filtered estimate of final GNP conditional on having the 75-day provisional GNP estimate, and its corresponding forecast error.

V. Estimation Results

The results of estimating equations (14a)-(14e) are shown in table 3. Almost uniformly, with one important exception discussed later, the estimated coefficients in table 3 are statistically different from zero at the 95 percent confidence level. Perhaps more importantly, again with the same exception just noted, the restrictions implied by equations (14) are all met. Thus, for example, equation (14a) implies that the sum of coefficients on the time-series forecast and the 15-day provisional estimate sum to one. In other words, the 15-day filtered estimate of real GNP growth is a simple weighted average of the forecast and 15-day GNP number. The last column reports the F-test statistic, and it clearly cannot reject the hypothesis of the 95 percent confidence level that the coefficients sum to unity.

Similarly, the restrictions in equations (14b) and (14d) that the coefficients sum to unity and that the coefficients on the lagged dependent variables equal the estimated correlation coefficients from table 2 are also met. In the case of equation (14d) the coefficient on $\hat{X}_{t/2}^*$ was not itself statistically significant, even though the joint hypothesis could not be rejected. When the coefficient on $X_{t/2}$ was constrained to be 0.9322—its a priori value as indicated by table 2—the coefficient on $\hat{X}_{t/2}^*$ became significant, which is the result reported in (14d).

Only equation (14e) gave any significant trouble. In this case, the estimated K_3 was not significantly different from zero, indicating that the 75-day estimate did not have any additional information about the final GNP number that was not already contained in the two preceding provisional estimates and the time-series forecast.

This last result stands in sharp contrast to the information provided by the first two provisional GNP numbers about final GNP. The estimated Kalman gain K_1 and K_2 in (14a) and (14c) are

TABLE 3

Estimated Filtering Equations 1974:IIQ - 1984:IQ (T-statistics in parentheses)

Equation	Standard Error of Estimate	F-Statistic
(14a) $X^*_t = 0.336 + 0.291\,\overline{X}_{t/0} + 0.774X_{t/1} + z_{1t}$ (0.89) (2.61) (8.45)	1.823	1.479[a]
(14b) $X_{t/2} = -0.047 + 0.816X_{t/1} + 0.236\,\hat{X}^*_{t/1} + v_{2t}$ (-.29) (7.33) (1.99)	0.714	2.247[b]
(14c) $X^*_t = -0.077 + 1.020\,X_{t/1} + 1.483\,v_{2t} + z_{2t}$ (-0.28) (19.61) (4.30)	1.477	2.378[c]
(14d) $X_{t/3} = 0.235 + 0.932X_{t/2} + 0.098\,\hat{X}^*_{t/2} + v_{3t}$ (2.54) (5.95)	0.560	0.578[d]
(14e) $X^*_t = 0.083 + 0.974\,\hat{X}_{t/2} + 0.784\,v_{3t} + z_{3t}$ (0.04) (21.86) (1.77)	1.361	0.339[e]

Addendum: time-series forecasting model

$X^*_t = 0.511 + 0.828\,X_{t-1} + w_t - 0.415w_{t-1}$
(2.19) (7.87) (2.40)

Standard error of estimate = 3.323

a. Test that sum of coefficients is unity.
b. Test that coefficient on $X_{t/1} = 0.0784$ and that sum of coefficients is unity.
c. Test that coefficient on $X_{t/2}$ is unity and that $K_2 = 1.234$.
d. Test that coefficient on $X_{t/2}$ is 0.932 and that coefficients sum to unity. The equation was reestimated with coefficient on $X_{t/2}$ restricted to 0.932; the results are the ones reported in (14d).
e. Test that coefficient on $\hat{X}^*_{t/2}$ is unity: the restriction that the coefficient was unity and that the coefficient on v_{3t} was nonzero was rejected at the 5 percent confidence level.
SOURCE: Author.

0.774 and 1.483, respectively, and both are statistically different from zero at the 95-percent confidence level. The latter number may seem too high—presumably it should be between zero and one. However, with serial correlation in the measurement errors the constraint is that $a_{23} K_2$ must be less than one. This constraint is satisfied by the calculated theoretical Kalman gains shown in table 4.[4] Clearly, the estimated K_2 of 1.483 is not statistically different from the theoretical value of 1.234, given the size of its standard error, a conclusion that is substantiated by the results of the F-test in table 3.

The nonzero coefficient on the time-series forecast variable, $\overline{X}_{t/0}$, suggests that the provisional estimates do not fully incorporate information about final GNP contained in the previous quarter's estimates.[5] This suggests perhaps a tendency on the part of the BEA to be conservative in extrapolating trends in the GNP data. And it also suggests that the typical practice in forecasting and policy analysis to discard forecasts for the immediately prior quarter once the provisional estimates become available may be inefficient.

The fact that the 75-day estimate does not appear to add any additional information about final GNP is interesting given that it is the first estimate to incorporate quarterly data. The high degree of serial correlation between the 45-day and 75-day provisional estimates shown in (14d), with relatively low variance in the residual, indicates, however, that the two estimates are not very different from each other despite the addition of the quarterly information. Indeed, to an important extent this is true of all three provisional estimates: they provide more information about each other than they do about the final GNP number.

■ 4 The similarity of the "theoretical" and estimated Kalman gains suggest there would be no advantage from calculating the filtered estimator using the theoretical numbers. There does not appear to be any clear consensus whether regression-based weighting of forecasts is preferable to sample-estimated optimal-weighting. See, for example, Lupoletti and Webb (1986), pp. 279-281.

■ 5 The time-series forecasts used only past data available or the time the new provisional estimator first became available, not past values of the final GNP growth estimator.

TABLE 4

Calculated and Estimated Kalman Gains

	Calculated	Estimated	Standard Error of Estimated Gains
K_1	0.719	0.774	(0.092)
K_2	1.234	1.483	(0.345)
K_3	0.679	0.784	(0.443)

SOURCE: Author.

VI. Conclusion

A recent and interesting analysis of the early GNP estimates has concluded that "they behave neither as efficient forecasts nor as observations measured with error" (Mork [1987], p. 173). The purpose of this paper has been to filter the early GNP numbers, to remove the measurement error, and to produce more accurate predictions of the final GNP growth estimates. In a related paper (Scadding [1987]), I have shown that these filtered estimates do not exhibit the unconditional bias and inefficiency that Mork found for the raw estimates. Another interesting sidelight of the results of this paper is that the Mankiw-Shapiro test for discriminating between observation and forecast errors does a poor job when applied to the estimated observation and forecast errors calculated in this paper, corroborating other indications of the poor power of the test.

For the forecaster, the filtering approach outlined in this paper provides an easy and systematic way of adjusting the provisional numbers to make them better estimates of "actual" GNP growth. It would be intriguing to inquire whether forecasters do in fact adjust the early numbers in a way that is consistent with the approach taken here.

The estimation results reported are model specific in the sense that they depend, to an unknown extent, on the specific forecasting model used to initialize the filtering procedure. Again, it would be interesting to see the extent to which the filtering results were sensitive to the forecasting model by using forecasts from alternative models. One offshoot of such an exercise would be that a particular model's performance could be evaluated in terms of the extent to which its forecasts contributed to improving the forecasting ability of the preliminary GNP numbers.

References

Bryson, Arthur Earl, and Yu-Chi Ho. *Applied Optimal Control.* Waltham, Mass.: Blaisdell Publishing Co., 1969.

Carson, Carol S. "GNP: An Overview of Source Data and Estimating Methods." *Survey of Current Business.* 67, no. 7 (July 1987): 103-26.

Conrad, William, and Carol Corrado. "Application of the Kalman Filter to Revisions in Monthly Retail Sales Estimates." *Journal of Economic Dynamics and Control.* 1, no. 2 (May 1979): 177-98.

Howrey, E. Philip. "Data Revision, Reconstruction and Prediction: An Application to Inventory Investment." *Review of Economics and Statistics.* 66, no. 3 (August 1984): 386-93.

Lupoletti, William M., and Roy H. Webb. "Defining and Improving the Accuracy of Macroeconomic Forecasts: Contributions from a VAR Model." *Journal of Business.* 59 (April 1986) part 1: 263-85.

Mankiw, N. Gregory, and Matthew D. Shapiro. "News or Noise? An Analysis of GNP Revisions." *Survey of Current Business.* 66, no. 5 (May 1986): 20-25.

McNees, Stephen K. "Estimating GNP: The Trade-off between Timeliness and Accuracy." *New England Economic Review.* Jan/Feb. 1986: 3-10.

Mork, Knut A. "'Ain't Behaving': Forecast Errors and Measurement Errors in Early GNP Estimates." *Journal of Business and Economics Statistics.* 5, no. 2 (April 1987): 165-75.

Scadding, John L. "The Nature of GNP Revisions." Federal Reserve Bank of Cleveland. *Working Paper* No. 8718, December 1987.

Walsh, C.B. "Revisions in the 'Flash' Estimate of GNP Growth: Measurement Error or Forecast Error?" Federal Reserve Bank of San Francisco. *Economic Review* (Fall 1985): 5-13.

Young, Allan H. "Evaluation of the GNP Estimates." *Survey of Current Business.* 67, no. 8 (August 1987): 18-42.

Comment—Intervention and the Dollar's Decline

by Owen F. Humpage

Owen F. Humpage is an economic advisor at the Federal Reserve Bank of Cleveland. This comment concerns an article he wrote for the preceding issue of *Economic Review* (Quarter 2 1988), pp. 2-16.

After publication of "Intervention and the Dollar's Decline" in the preceding issue of *Economic Review,* some confusion arose regarding exactly when the exchange-rate quotes in that article were taken and from what market they were derived. This comment will explain the differences and respecify some of the equations to dispel any misinterpretation.

The daily data for the article were taken from DRI-FACS in August 1987. We understood from reading the DRI-FACS manual that the data series from August 7, 1984 to August 28, 1987 were morning opening exchange-rate quotes from the New York market.

The recently revised DRI-FACS manual (now called DRIFACS PLUS) indicates that after October 8, 1986, the data refer to closing quotes in the London market.[1] We therefore reestimated the equations in tables 3 and 4 of the article to determine if this change had any significant effect on the results.

While some of the point estimates are slightly different under these new estimations, the overall conclusion of the article remains the same: Between August 1984 and August 1987, day-to-day U.S. intervention did not systematically affect day-to-day exchange-rate movements. However, on some occasions, intervention did have a temporary effect on mark-dollar and/or yen-dollar exchange rates.

Statistical tests in the article included U.S. intervention with a one-day lag to avoid problems with bidirectional causality between exchange rates and intervention. Generally, the results are interpreted on the assumption that the effects of U.S. intervention on day t-1 occurred between the opening quote on day t-1 and the opening quote on day t. After October 8, 1986, however, the data are closing quotes from the London market. Since the New York market opened before the London market closed, U.S. intervention on day t-1 could have affected the London closing exchange-rate quote on day t-1 and on day t.

To allow for this possibility, we reestimated the relevant equations, including a contemporaneous intervention term. Tables 3A and 4A, which correspond to tables 3 and 4 of the original article, present the results.

■ 1 DRIFACS PLUS, the Dictionary of Money Markets and Fixed Income Data, Data Resources, Inc., February 1988. Data prior to October 8, 1986 are as originally reported.

TABLE 3A

G7 Intervention

I. Estimation Period: February 23, 1987 to July 2, 1987

A. Dependent Variable: mark-dollar exchange rate

Independent Variables			Coefficient	T-statistic
Intervention dummies				
Initial purchases	no lag	(1)	0.009	1.73[a]
	lagged	(1)	-0.007	-1.35[b]
Subsequent purchases	no lag	(0)	—	—
	lagged	(0)	—	—
Initial sales	no lag	(3)	-0.007	-2.38[c]
	lagged	(3)	-0.006	-2.06[c]
Subsequent sales	no lag	(2)	-0.006	-1.14
	lagged	(2)	-0.008	-1.56
Lagged dependent			1.00	994.8[d]

Sum of Squared Residuals = 0.002
R^2 = 0.824
n = 90

B. Dependent Variable: yen-dollar exchange rate

Independent Variables			Coefficient	T-statistic
Intervention dummies				
Initial purchases	no lag	(0)	—	—
	lagged	(0)	—	—
Subsequent purchases	no lag	(0)	—	—
	lagged	(0)	—	—
Initial sales	no lag	(2)	-0.011	-1.89[a]
	lagged	(2)	-0.001	-0.21
Subsequent sales	no lag	(16)	-0.007	-3.08[d]
	lagged	(16)	0.0005	0.21
Lagged dependent			1.000	7016.4[d]

Sum of Squared Residuals = 0.003
R^2 = 0.969
n = 90

NOTE: Intervention refers to U.S. purchases or sales of foreign currencies. Numbers in parentheses indicate the number of times the dummy equals 1.
a. Significant at the 10% confidence level.
b. Significant at the 10% confidence level (one-tailed).
c. Significant at the 5% confidence level.
d. Significant at the 1% confidence level.
SOURCE: Author's calculations.

Table 3A lists the results for the period February 23, 1987 to July 2, 1987. For the West German mark, the coefficient for initial purchases of marks is positive and significant. One cannot interpret this coefficient unambiguously, because causality is bidirectional without the lag; nevertheless, the positive cofficient is not consistent with the view that intervention purchases of marks produced a dollar depreciation.

The lagged value on initial intervention is marginally significant and correctly signed. The United States bought a small amount of marks on March 11, as the dollar rose above 1.85 marks. The dollar depreciated on the following day. The coefficients on the sales of marks are incorrectly signed and/or insignificant. For the Japanese yen, all of the coefficients are either incorrectly signed or insignificant.

Table 4A presents the results for the period July 5, 1987 to August 28, 1987. For the West German mark, the coefficient for initial purchases of marks is positive and significant. As before, this coefficient cannot be unambiguously interpreted, but the sign is not consistent with the view that intervention purchases of marks produced a dollar depreciation. The remaining intervention variables are not significant. For the yen, the coefficients are either incorrectly signed or are not significant.

TABLE 4A

G7 Intervention

I. Estimation Period: July 5, 1987 to August 28, 1987

A. Dependent Variable: mark-dollar exchange rate

Independent Variables			Coefficient	T-statistic
Intervention dummies				
Initial purchases	no lag	(1)	0.012	2.53[a]
	lagged	(1)	–0.001	–0.27
Subsequent purchases	no lag	(3)	0.003	0.75
	lagged	(3)	0.002	0.47
Initial sales	no lag	(0)	—	—
	lagged	(0)	—	—
Subsequent sales	no lag	(0)	—	—
	lagged	(0)	—	—
Lagged dependent			0.999	758.5[b]

Sum of Squared Residuals = 0.001
R^2 = 0.849
n = 38

B. Dependent Variable: yen-dollar exchange rate

Independent Variables			Coefficient	T-statistic
Intervention dummies				
Initial purchases	no lag	(0)	—	—
	lagged	(0)	—	—
Subsequent purchases	no lag	(0)	—	—
	lagged	(0)	—	—
Initial sales	no lag	(1)	–0.018	–2.51[a]
	lagged	(1)	0.009	1.20
Subsequent sales	no lag	(0)	—	—
	lagged	(0)	—	—
Lagged dependent			1.000	4166.2[b]

Sum of Squared Residuals = 0.002
R^2 = 0.830
n = 38

NOTE: Intervention refers to U.S. purchases or sales of foreign currencies. Numbers in parentheses indicate the number of times the dummy equals 1.
a. Significant at the 5% confidence level.
b. Significant at the 1% confidence level.
SOURCE: Author's calculations.

Economic Review

■ Quarter III 1987

Can Services Be a Source of Export-led Growth? Evidence from the Fourth District
by Erica L. Groshen

Identifying Amenity and Productivity Cities Using Wage and Rent Differentials
by Patricia E. Beeson and Randall W. Eberts

FSLIC Forbearances to Stockholders and the Value of Savings and Loan Shares
by James B. Thomson

■ Quarter IV 1987

Learning, Rationality, the Stability of Equilibrium and Macroeconomics
by John B. Carlson

Airline Hubs: A Study of Determining Factors and Effects
by Paul W. Bauer

A Comparison of Risk-Based Capital and Risk-Based Deposit Insurance
by Robert B. Avery and Terrence M. Belton

■ Quarter I 1988

Can Competition Among Local Governments Constrain Government Spending?
by Randall W. Eberts and Timothy J. Gronberg

Exit Barriers in the Steel Industry
by Mary E. Deily

Why Do Wages Vary Among Employers?
by Erica L. Groshen

■ Quarter II 1988

Intervention and the Dollar's Decline
by Owen F. Humpage

Using Financial Data to Identify Changes in Bank Condition
by Gary Whalen and James B. Thomson

Developing Country Lending and Current Banking Conditions
by Walker F. Todd

Third Quarter Working Papers

Working Paper Notice

Current *Working Papers* of the Cleveland Federal Reserve Bank are listed in each quarterly issue of the *Economic Review*. Copies of specific papers may be requested by completing and mailing the attached form below.

Single copies of individual papers will be sent free of charge to those who request them. A mailing list service for personal subscribers, however, is not available.

Institutional subscribers, such as libraries and other organizations, will be placed on a mailing list upon request and will automatically receive *Working Papers* as they are published.

■ **8805**
Lessons of the Past and Prospects for the Future in Lender of Last Resort Theory
by Walker F. Todd

■ **8806**
Capital Requirements and Optimal Bank Portfolios: A Reexamination
by William P. Osterberg and James B. Thomson

■ **8807**
Testing for Speculative Bubbles in Stock Prices
by Asli Demirguc-Kunt and Hashem Dezhbaksh

■ **8808**
Financial Structure and the Adjustment of Capital Stock
by William P. Osterberg

■ **8809**
Interurban Comparisons of the Quality of Life
by Patricia E. Beeson and Randall W. Eberts

■ **8810**
Using SMVAM as a Linear Approximation to a Nonlinear Function: A Note
by Asli Demirguc-Kunt and James B. Thomson

Please complete and detach the form below and mail to:

Federal Reserve Bank of Cleveland
Research Department
P.O. Box 6387
Cleveland, Ohio 44101

Check item(s) requested.

Please send the following Working Paper(s):

☐ **8805** ☐ **8807** ☐ **8809**

☐ **8806** ☐ **8808** ☐ **8810**

Send to:

Please print

Name

Address

City State Zip