Economic Review

Federal Reserve Bank of San Francisco

1993 Number 3

Table of Contents

The Federal Reserve Bank of San Francisco's Economic Review is published three times a year by the Bank's Research Department under the supervision of Jack H. Beebe, Senior Vice President and Director of Research. The publication is edited by Judith Goff. Design, production, and distribution are handled by the Public Information Department, with the assistance of Karen Flamme and William Rosenthal.

For free copies of this and other Federal Reserve publicatons, write or phone the Public Information Department, Federal Reserve Bank of San Francisco, P.O. Box 7702, San Francisco, California 94120. Phone (415) 974-2163.

Printed on recycled paper with soybean inks.

Using a Nominal GDP Rule to Guide Discretionary Monetary Policy

John P. Judd and Brian Motley

The authors are Vice President and Associate Director of Research and Senior Economist, respectively, at the Federal Reserve Bank of San Francisco. They would like to thank Jack Beebe, Chan Huh, Bennett McCallum, Ann-Marie Meulendyke, Glenn Rudebusch, John Taylor, Bharat Trehan, and Carl Walsh for helpful suggestions, and Andrew Biehl for excellent research assistance. Any remaining errors are the sole responsibility of the authors.

Given doubts about the reliability of the monetary aggregates as intermediate targets of monetary policy, the Federal Reserve attempts to meet its dual goals—gradual reduction of inflation and mitigation of cyclical downturns in output—through purely discretionary adjustments of an interest rate instrument in response to myriad incoming data. A procedure in which the Fed would consult a nominal GDP feedback rule, while retaining the flexibility to use discretion in its monetary policy decisions, might contribute to achieving its long-run inflation goal without significantly interfering with its ability to pursue its short-run cyclical goal. This paper describes such a policy regime, and presents some empirical evidence pertinent to an assessment of how it might work.

In recent years, the Federal Reserve has become more explicit about its desire to reduce and ultimately eliminate inflation, citing the beneficial effects of stable prices on long-term economic growth.[1] At the same time, it has retained the goal of mitigating cyclical downturns in employment and output. Dual goals inevitably raise the issue of which should take precedence when they have conflicting implications for policy. The Fed resolves these conflicts on a case-by-case basis, using its discretion to set policy after analyzing a wide array of real and financial indicators. Like most of its counterparts in other industrial countries, it uses a short-term nominal interest rate (the federal funds rate) as its policy instrument (Kasman 1993).

In pursuing gradual disinflation over the last fifteen years, the Fed has attempted to use the monetary aggregates (mainly M2 since 1982) as a "nominal anchor" to help prevent short-term discretionary decisions from inadvertently allowing inflation to stray from the long-term goal. Annual target ranges have been established for various measures of money, and these have been lowered gradually over time to be consistent with declining inflation. The idea underlying this approach is that if the funds rate were adjusted so that money fell within these declining ranges over time, inflation would be slowed correspondingly. At the same time, the ranges are wide enough to permit flexibility to respond to cyclical downturns.

However, the Fed has de-emphasized the monetary aggregates because their relationships with prices and output have deteriorated, apparently in response to financial deregulation and innovation (Judd and Trehan 1992). As a consequence, policy has been left without much guidance from a nominal quantity variable that is closely linked to inflation in the long-run.

This void makes it difficult to tell if short-run decisions about the funds rate are consistent with the long-run objective of lowering inflation. Moreover, since there is no single variable that provides an *automatic* signal to policymakers that an interest rate change should be seriously contemplated, each change in the funds rate must be made on a judgmental, case-by-case basis. Perhaps inevitably, there is a temptation for policymakers to respond more

1. See Judd and Beebe, this issue. For a discussion of the possible benefits of low inflation, see Howitt (1990).

strongly and quickly to shocks that threaten a recession than to those that are stimulative, and this can test policymakers' resolve to control inflation (Barro 1986).

Use of an interest rate instrument without the guidance of a nominal anchor also tends to foster the questionable view that the stance of policy can be characterized by the level of the funds rate. As a result, tightening or easing policy becomes defined as raising or lowering the funds rate, while a decision to leave the funds rate unchanged is seen as no change in policy. However, these characterizations can be misleading. For example, a constant interest rate can be consistent with either tighter or easier policy, depending upon what is happening to the other determinants of aggregate demand. Thus a sudden rise in consumer confidence that leads to less saving could render a constant interest rate more expansionary. The same problem also arises for long-run inflation policy. To maintain a constant policy with respect to the inflation rate, interest rates would need to change frequently to offset the effects of shocks and assure that aggregate demand grew in line with the economy's productive potential.

The purpose of this paper is to describe a monetary policy regime in which *discretionary* changes in a short-term interest rate would be oriented around a *baseline* interest rate path that would be designed to be consistent with a disinflation or low-inflation goal. Specifically, under this approach, the baseline (or no-change-in-policy) option would be defined by a policy rule that would link changes in a short-term interest rate to a nominal GDP target designed to be consistent with the inflation goal in the long run. Thus, the rule would provide information to policymakers in formulating short-term discretionary actions that might help them avoid inadvertently allowing inflation to drift away from its desired level over time. The nominal GDP target either could be made public or used for internal purposes only.

The remainder of this paper is organized as follows. Section I discusses the relationship between monetary and nominal GDP targets, and argues that the latter have intrinsic appeal when unstable velocity makes monetary targets unreliable. In Section II, simulations of a specific nominal GDP rule are presented as an example to illustrate some properties of such rules. Section III discusses how a nominal GDP rule could be used to inform a discretionary monetary policy, and concludes by briefly noting some practical problems that would need to be solved in actually implementing such an approach.

I. NOMINAL GDP TARGETS

In this section, we discuss why nominal GDP may have some appeal as an intermediate target of monetary policy, especially as an alternative to the monetary aggregates when their velocities become unstable.

The channel of influence from nominal GDP growth to inflation can be seen from the following definition, which states that inflation is equal to the difference between growth in nominal and real GDP:

$$\Delta p \equiv \Delta x - \Delta y, \tag{1}$$

where Δp, Δx, and Δy represent the annualized growth rates of the implicit GDP deflator, nominal GDP, and real GDP, respectively. In the long-run, real GDP growth can be approximated by a trend rate that is determined by real factors including the growth in labor, capital, and productivity, and thus is largely independent of nominal GDP growth.[2] As a consequence, any given growth rate of nominal GDP can be translated into a corresponding inflation rate in a straightforward way.[3] For example, trend (or potential) real GDP growth commonly is estimated at around 2 percent, so that a 5 percent growth rate of nominal GDP would fix long-run inflation at around 3 percent.

Since the growth rate of nominal GDP is equal to the growth rate of money (Δm) plus the growth rate of velocity (Δv), targeting money can be seen as an indirect method of targeting nominal GDP. Thus,

$$\Delta x \equiv \Delta m + \Delta v, \tag{2}$$

Putting these definitions together yields,[4]

$$\Delta p \equiv \Delta m + \Delta v - \Delta y. \tag{3}$$

So long as trend velocity growth is stable, any given long-run growth rate of money can be translated into a long-run inflation rate in a straightforward manner. When the velocity of M2 was stable, the relationship between M2 and inflation was particularly simple, since historically the

2. This statement abstracts from possible effects of trend inflation on trend real GDP growth. Thus lower (higher) nominal GDP growth will result in lower (higher) inflation, which for various reasons may be associated with higher (lower) trend growth of real GDP (Motley 1993). However, these effects are likely to be small when compared with the range of nominal GDP growth rates and inflation observed in the past.

3. We have specified the nominal GDP identity in terms of growth rates rather than levels. Research suggests that the steady-state growth rate of real GDP is stationary, so that fixing the growth rate of nominal GDP will result in a stationary inflation rate. The situation is more complex when the equation is specified in levels. It is uncertain whether the level of real GDP is stationary or not, so that it is difficult to tell if the price level would be stationary under a nominal GDP level target.

4. The relationships discussed in this paragraph are growth-rate versions of the ones behind the P^* model (Hallman, Porter and Small, 1991).

trend growth rate of M2 velocity was zero. Thus, for example, a 5 percent growth rate of M2 would produce 5 percent nominal GDP growth and 3 percent inflation in the long run. However, *when velocity is unstable,* direct nominal GDP targeting has the advantage that it is not adversely affected by unpredictable swings in velocity. In effect, nominal GDP targeting is a way to circumvent problems with the velocity of money in conducting monetary policy.[5]

The principal drawback to using nominal GDP as an intermediate target is that it does not respond as promptly as money does to the Fed's policy instruments, and hence is not very controllable, even over periods as long as several quarters. Thus it would be difficult for the Fed, or the public, to know if day-to-day policy actions were consistent with achieving the nominal GDP target over time. One way of dealing with this control problem is to compare discretionary policy changes to those called for by a *feedback rule,* which specifies responses of the policy instrument to incoming data on nominal GDP.[6]

A feedback rule of the type suggested by McCallum (1990), for example, would specify that the policy instrument would be adjusted in each period by a predetermined proportion of the difference between actual and targeted nominal GDP in the prior period. If the instrument were set strictly according to a properly specified rule of this type, the nominal GDP target would be achieved to a reasonable approximation over the long run, even though it might be missed over shorter time periods. Hence, a practice of orienting *discretionary* changes in the policy instrument around such a baseline would provide policymakers with information they could use to help them achieve their nominal GDP target over the long haul. And, achieving the nominal GDP target in the long run would hold average inflation to within a reasonable range around its target.

II. EXAMPLE: A NOMINAL GDP GROWTH RATE RULE

A number of different nominal GDP feedback rules have been explored in the literature. These differ as to whether the policy instrument is a reserves aggregate or a short-term interest rate, and whether nominal GDP and/or the policy instrument are specified in levels or growth rates. A common feature of these feedback rules is that the Fed responds to actual data on nominal GDP rather than to forecasts.[7] This feature has an advantage when decisions are being made by a committee of individuals who may disagree about the implications of incoming data for the future path of nominal GDP.

Below we briefly review research on a policy regime in which the Fed changes the short-term interest rate in response to divergences between actual and targeted nominal GDP *growth rates* (Judd and Motley 1992).[8] A short-term interest rate is specified as the instrument because it is controllable in the short run and because the FOMC has shown a preference over the years for operating through such an instrument rather than a reserves aggregate.[9] Our purpose in presenting this example is neither to advocate this particular form of feedback rule, nor to advocate strict adherence to any rule. Rather our purpose is to show how this class of rules might work as a baseline for discretionary policy decisions. The rule we have examined is specified as follows:

$$\Delta R_t = \lambda(\Delta x_{t-1} - \Delta x^*_{t-1}),\ \lambda > 0. \tag{4}$$

In this equation, ΔR_t is the quarterly percentage point change in a short-term interest rate (we used the three-month Treasury bill rate), and Δx_{t-1} and Δx^*_{t-1} are the

5. Given that the ultimate objective of long-run monetary policy is to control inflation, it might make sense to target the rate of inflation directly. However, as shown in Judd and Motley (1991), the lags from monetary policy to the rate of inflation appear to be sufficiently long in Keynesian-type (sticky-price) models that attempts at direct inflation targeting might result in extreme volatility in the interest rate and real GDP. Since it is desirable to select a rule that is robust across alternative types of models, we have not focused on direct inflation targeting in this paper.

6. The feedback rule discussed later in this paper is specified along the lines of rules originally proposed and analyzed by McCallum (e.g., see his 1990 paper). Feedback rules also have been examined by a number of other researchers, including Hess, Small and Brayton (1993), Judd and Motley (1991, 1992), Meltzer (1987), and Taylor (1985, 1992).

7. This feature of the rule could be modified to incorporate more up-to-date information by replacing last quarter's nominal GDP growth rate with a projection of the current quarter's data. Such short-term forecasts would be purely a matter of interpreting monthly indicator variables and would not depend very much upon views of the structure of the economy. As such they would not violate the spirit of the feedback rule.

8. Specifying a rule in terms of the change, rather than the level, of the interest rate has the advantage that it is not necessary to know in advance the equilibrium level of the real interest rate. Under a properly specified rule for the change in the nominal interest rate, the economy automatically would tend to adjust such that the real interest rate would move toward its equilibrium level over time, whatever that level happened to be.

9. As shown in Judd-Motley (1992), in principle, a reserves aggregate offers the possibility of much tighter control over inflation than appears likely under an interest rate instrument. The main difficulty with these aggregates as instruments of policy is that financial innovation and deregulation have made the velocities of reserves and the monetary base highly unstable. Moreover, increased international (paper) currency movements have added to problems with the velocity and controllability of the monetary base.

actual and targeted annualized growth rates of nominal GDP in the preceding quarter.[10]

The nominal GDP growth rate target would be chosen to be consistent with the target for inflation. For example, a goal of reducing inflation gradually to zero and holding it there would imply that the nominal GDP growth target would be lowered gradually toward 2 percent and held at about that pace.

The strength of the interest rate response to a given target deviation is defined by λ, and would be chosen by the central bank. As discussed below, simulations suggest that a value of λ of 0.2 would be sufficient to achieve reasonable control of inflation, without raising the volatility of output or interest rates compared with actual experience in the past three decades. This value of λ implies that the interest rate would be raised (lowered) by 20 basis points during each quarter in which the annualized nominal GDP growth rate exceeded (fell short of) the target by 1 percentage point. Although this may seem a rather weak response, it is important to recognize that under the rule the interest rate would continue to be raised (lowered) each quarter so long as growth remained above (below) target. According to the simulations, the consistent application of this modest response is sufficient to hold nominal GDP growth near its target over the long haul.

Simulation Results

In order to obtain a rough idea about how implementation of this rule might affect the economy, we employed simulations of two simple macroeconomic models under the assumption that the rule was in place and the economy was hit by shocks like those that actually occurred. We did large numbers of stochastic simulations so that we could construct confidence intervals for the outcomes for inflation, real GDP, and the short-term interest rate. In constructing these simulations, we had to assume that the rule was followed precisely. If the rule were used as a baseline for discretionary policy, the policymakers could attempt to improve on these results in whatever ways they deemed appropriate.

As with any counterfactual simulations, these exercises are subject to some valid criticisms, which mean that such results should be interpreted with caution. First, the simulation results will depend upon the particular model(s) used. Since individuals will differ as to what they think characterizes a reasonable model, simulation results may be suspect. In an attempt to deal with this problem, the simulations were run with two alternative models, a small Keynesian model, and a (largely) atheoretical vector error-correction model.

Second, counterfactual simulations are subject to the Lucas critique that the structure of the economy would have been different from history if the rule actually had been used. To attempt to deal with this concern, we varied the key coefficients in the models and re-ran the simulations to test for robustness. As discussed in Judd-Motley (1992), based upon these exercises, we concluded that the results were not particularly sensitive to the alternatives considered, although there were some instances in which coefficient changes did significantly affect the simulation results. We do not consider our study, or any other single study, to be definitive, and it would be useful to test this and other rules further in the context of other models.

The simulations suggest that following such a rule would have provided for improved control of inflation compared with actual experience over the past three decades. As measured by the GDP deflator, actual inflation averaged 5½ percent over the 30-year sample. The simulation results suggest that average annual inflation over 1960–1989 would have been held to between about zero and about 2½ percent (depending on the model) with a probability of two-thirds (see the box.) Moreover, it appears that this result could have been achieved without significantly increasing the volatility of real GDP and with a reduction in the volatility of interest rates compared to historical experience.[11] The lessened interest rate swings

10. As an alternative, the rule could specify a target for *average* nominal GDP growth over more than one prior quarter. For example, the nominal GDP target could be specified each quarter in terms of growth over the prior half year or full year. This approach would have the advantage of smoothing out quarter-to-quarter volatility in nominal GDP growth (whether due to "noise" in preliminary data, inventory cycles, or other factors) that might otherwise induce unnecessary interest rate responses. The disadvantage of using averages of several past quarters of nominal GDP growth would be that it introduces additional lags into the interest rate responses under the rule. Simulation experiments with the models referred to in this paper suggest that these longer lags tend to increase the size of cycles in real GDP and inflation that might occur under a mechanical application of the rule. In effect, using an average of several prior quarters of nominal GDP growth delays the response of interest rates to deviations of nominal GDP from the target, and thus tends to set off cycles of overshooting followed by undershooting of that target.

11. According to our simulations, a rule that focuses on the *growth rate,* rather than the *level,* of nominal GDP has the advantage of producing less volatility in real GDP and interest rates. However, the growth rate rule has the disadvantage that the price level could drift over time in the event of a prolonged series of positive or negative shocks. One way of attempting to deal with this problem would be to provide for occasional adjustments to the nominal GDP growth target when it permitted unacceptably large price-level (or nominal GDP) drift. This method might help to preserve the price level in the long run, while retaining the benefits of less volatility most of the time.

SIMULATIONS OF THE NOMINAL GDP GROWTH RATE RULE

Below we present results of simulations that assess how the macroeconomy might have evolved over the past three decades if the nominal GDP growth rate rule had been in use, and the structure of the economy had remained unchanged. In these "counterfactual simulations," the targeted values of nominal GDP growth were set to be consistent with zero inflation over 1960–1989. We used a value for λ of 0.20. For each of two models (a small Keynesian and a VECM, described in Judd and Motley 1992, pp. 14-16), we calculated 500 stochastic simulations, where the random shocks in each model equation were drawn from distributions that had the same means and variances as the estimated error terms.

We measure inflation performance in terms of average annual inflation over the simulation period. The volatility of output is measured in terms of the four-quarter growth rate of real GDP. Finally, the volatility of interest rates is measured as quarter-to-quarter changes in the three-month Treasury bill rate. The results of the simulations are shown below in the form of one standard deviation confidence bands (thus, two-thirds of the stochastic simulations fell within the bands.)

STOCHASTIC SIMULATIONS WITH NOMINAL GDP GROWTH RULE
ONE-STANDARD DEVIATION CONFIDENCE INTERVALS
1960-1989

	AVERAGE ANNUAL INFLATION RATE	4-QUARTER REAL GDP GROWTH RATE	QUARTER-TO-QUARTER CHANGE IN INTEREST RATE
ACTUAL	5.4%	0.5% to 5.5%	−1.0% to 1.0%
RULE			
Keynesian Model	−0.2% to 2.7%	1.2% to 5.8%	−0.6% to 0.9%
VECM	0.4% to 2.1%	2.5% to 7.5%[a]	−0.6% to 0.8%

[a]Taken literally, the results of the VECM simulations suggest that achieving lower inflation would produce an average rate of growth of real GDP that is above the experience in the U.S. in the post World War II period. This result reflects the well-known negative correlation observed in the U.S. data between inflation and real GDP growth, which is embedded in the VECM coefficients. This correlation could reflect the effects of inflation on growth, and/or the effects of supply shocks (e.g., oil shocks) on both variables. Since the VECM is not designed to distinguish between these two effects, our results should be interpreted as agnostic concerning the extent to which low inflation might boost long-term growth. In this paper, we take as given the Fed's stated goal of gradually moving toward price stability, and do not attempt to assess the possible effects of such a policy on average real GDP growth.

apparently arise because a consistent application of the rule keeps the inflation rate under control in the simulations, and thus highly aggressive policy responses are not likely to be needed. Thus, for example, use of the rule prevents simulated inflation from rising as sharply in the mid 1970s and early 1980s, and thus moderates the size of any policy tightening that might have been necessary to return inflation to lower levels.

One potential problem with these simulations is that they do not take into account the effects of measurement errors in nominal GDP. Even though the rule involves policy reactions to "actual" data on nominal GDP lagged one quarter, these data are revised a number of times before they are considered final. Measurement errors in the early releases of nominal GDP data, which policymakers would observe as they used the rule, would induce movements in interest rates and thus also affect outcomes for real GDP and inflation. In order to estimate the size of any such effects, we re-ran the above simulations with measurement errors (equal in size to those observed over 1978–1989) added to the "observations" of nominal GDP in the rule.[12]

12. The measurement errors were introduced as white-noise shocks with a standard deviation of 1.5 percent (annual rate), which is equal to the standard deviation of the differences between the "advance" nominal GDP growth rates and the "latest revised" nominal GDP growth rates over 1978-1989 (Bureau of Economic Analysis). Following the approach of Gagnon and Tryon (1993), the model was estimated with final revised data, but we added shocks representing measurement errors to the nominal GDP growth rates that enter the rule. Thus, the rule used in these simulations was: $\Delta R_t = \lambda\,(\Delta x_{t-1} + \epsilon_{t-1} - \Delta x^*_{t-1})$, where ϵ_{t-1} represents measurement error. We also investigated the possibility that the revisions are autocorrelated by estimating first and second order

This exercise yielded confidence intervals for inflation, real GDP and interest rates very close to those shown in the box—in fact, no confidence interval was increased in width by more than 0.1 percentage point.[13]

Comparison of the Rule with Actual Policy

How would the nominal GDP growth rate rule have performed in recent years in comparison with actual policy? To shed some light on this issue, we conducted counterfactual simulations over 1988–1993 in which we assumed that the economy was hit by the same set of shocks that actually occurred during this period. Consistent with the Fed's objective to lower the inflation rate gradually over time, we (somewhat arbitrarily) assumed targets for nominal GDP growth that declined by ¼ percent per year from 7 percent in 1988 to 5¾ percent in 1993, so that they roughly matched the *overall* decline of nominal GDP growth rates over the period.

As shown in Figure 1, the simulated path of the interest rate generated by this combination of target path and rule is fairly close to the path that actually occurred.[14] These simulations were computed using the latest revised data, rather than the data the FOMC actually observed at the time. When the measurement errors in these data are accounted for in the simulations, the short-term interest rate is about 50 basis points lower (in both models) over mid-1990 to mid-1993 than the simulation shown in Figure 1. In the final four quarters shown in Figure 1, the simulated interest rates with and without measurement errors bracket the actual level of the interest rate (for both models).

These simulations suggest that, even though the Fed was not following a nominal GDP rule during this period, actual policy was not inconsistent with that indicated by the rule in combination with a disinflationary path for nominal GDP.[15] It should be noted, however, that the level (but not the pattern) of the simulated interest rate is sensitive to the exact level of the assumed nominal GDP growth rate targets. Thus, for example, an equally plausible set of targets that consistently were ½ percentage point lower than the ones assumed would produce a simulated interest rate path that was parallel and uniformly higher than the one in the figure.

Figure 2 shows simulations of the interest rate paths that would be produced by adopting alternative targets for nominal GDP growth, and compares them with the path produced by the nominal GDP target assumed in Figure 1. The line marked "easy" corresponds to a nominal GDP growth rate target that remains unchanged at 7 percent in 1988 through mid 1993. The line marked "tight" simulates what might have happened if the nominal GDP growth rate target had been reduced by ½ percent per year from 7 percent in 1988 to 4½ percent in 1993. As can be seen, the constant nominal GDP growth rate target is projected to involve a lower interest rate by 1993 than projected under the gradual disinflation targets of Figure 1 (labeled "moderate"), while the more rapid ½-percent-per-year decline in the nominal GDP growth target under the "tight" policy would have involved a higher interest rate. Under all three target paths for nominal GDP growth, the interest rate would have fallen noticeably in the 1990–1991 recession.

III. Informing Discretionary Policy Decisions with a Rule

The above discussion of the nominal GDP growth rate rule was not designed to advocate that particular feedback rule as a baseline for a discretionary policy, but rather to provide a specific illustration of the properties of this class of rules. Given the demise of the monetary aggregates as reliable intermediate targets, the FOMC attempts to meet its dual goals (control of inflation and the mitigation of cyclical downturns in output) through purely discretionary adjustments of an interest rate instrument in response to myriad incoming data. A procedure in which the FOMC would *consult* a nominal GDP feedback rule, while retaining the flexibility to use discretion in its short-run decisions, might contribute to achieving its inflation goal without significantly interfering with its ability to pursue

autocorrelation coefficients of nominal GDP revisions over 1976-1983, as shown in Walsh (1985). Autocorrelation was rejected at very high marginal significance levels. Despite this result, we experimented with first-order autocorrelated revision errors (with standard error of 1.5 percent) in the simulations, and found that their effect was virtually the same as the white-noise errors as long as the autocorrelation coefficient was less than 1.

13. The small effect of measurement errors in the simulations results from several factors. First, the size of the typical revision to nominal GDP in recent years is sufficient to have only very modest effects on the short-term interest rate in the nominal GDP rule. For example, a one-standard deviation revision (1.5 percent, annual rate) induces a change in the interest rate of 30 basis points. Second, as is typical of macroeconomic models, the coefficients linking changes in interest rates to changes in real GDP and inflation in the models used in this paper are relatively small. Third, interest rates affect real GDP and inflation with relatively long distributed lags. Thus measurement errors of opposite signs will tend to have offsetting effects on real GDP and inflation.

14. Similar results were obtained when the simulation was begun in later years.

15. Taylor (1992) has obtained a similar result with a different nominal GDP rule, using data prior to the most recent re-benchmarking.

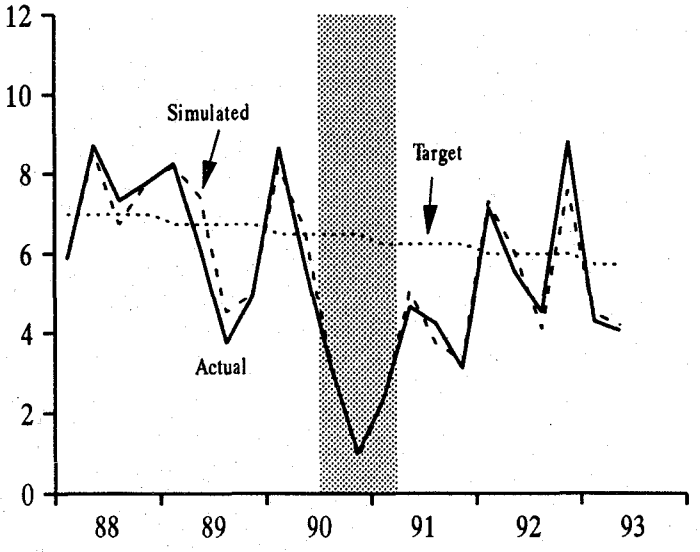
FIGURE 1

SIMULATIONS OF A POLICY RULE

KEYNESIAN MODEL

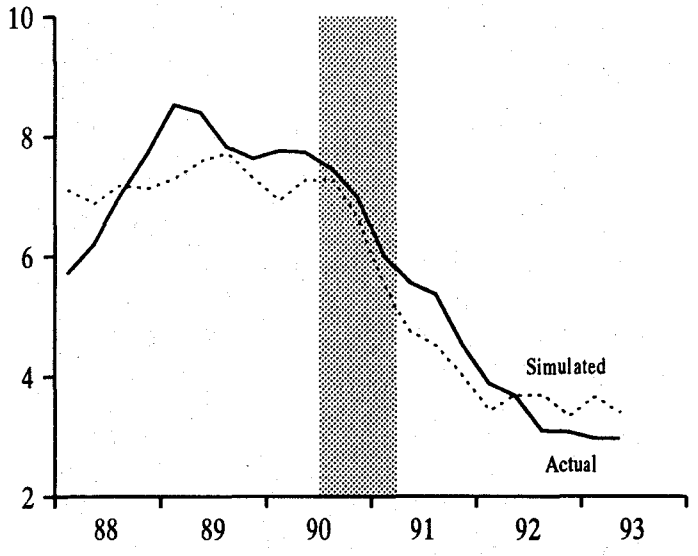
Nominal GDP Growth

Percent



3-Month T-Bill Rate

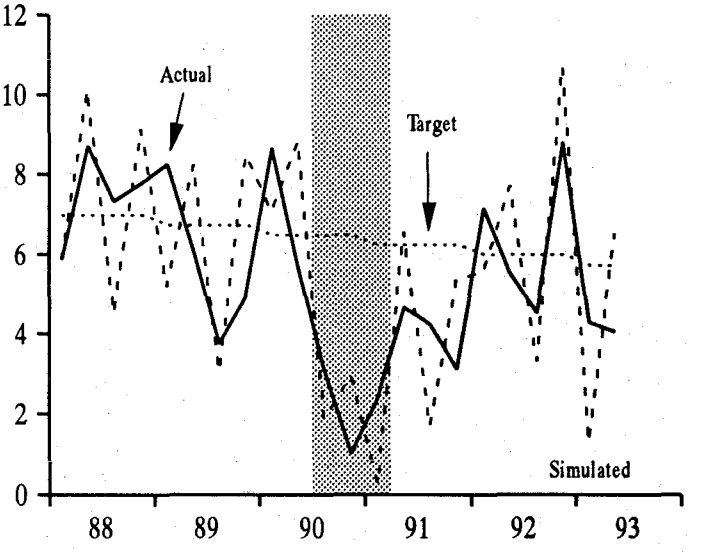
Percent



VECTOR ERROR CORRECTION MODEL

Nominal GDP Growth

Percent



3-Month T-Bill Rate

Percent

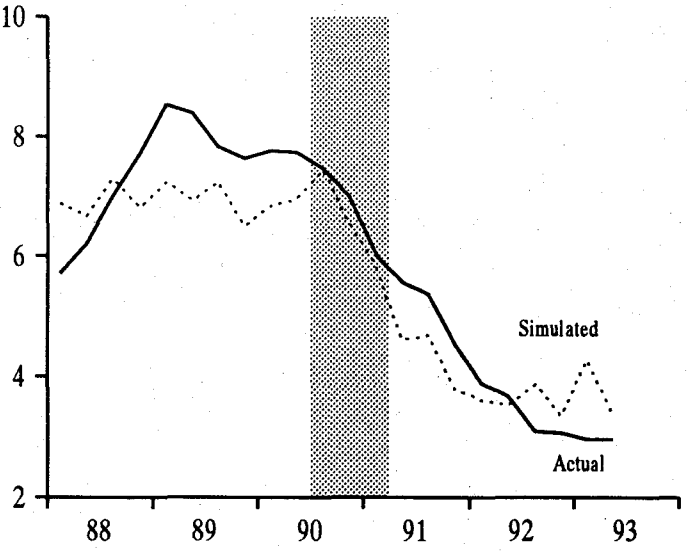
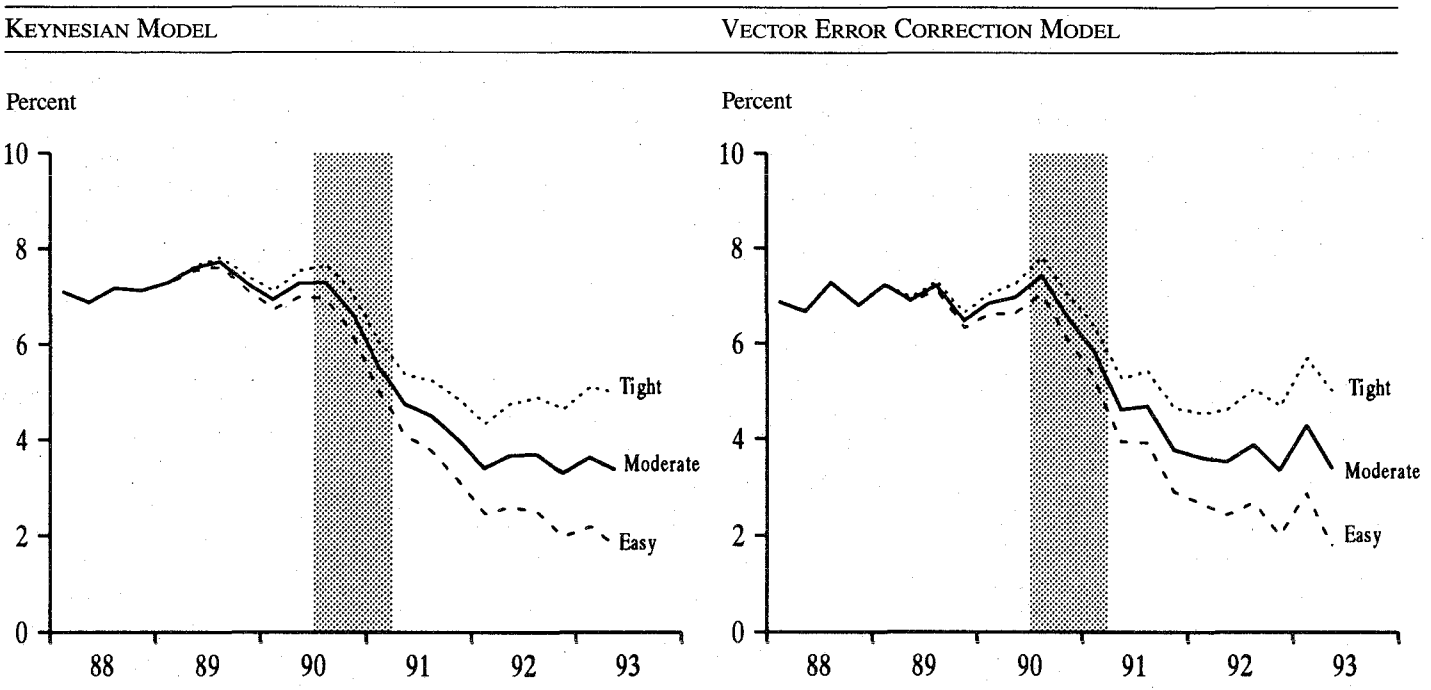


FIGURE 2

Alternative Policy Simulations: 3-Month T-Bill Rate



its cyclical goal. Such a rule could be announced to the public or used for internal information only.

Consultation with a rule could take several forms. For example, Taylor (1992) has suggested that the FOMC simply include the interest-rate "recommendations" of a nominal GDP feedback rule with any other monetary policy indicators they wish to consult.

While this idea seems reasonable, the decision-making process might benefit by having a feedback rule play a more central role. Specifically, the interest rate path indicated by a rule could be defined explicitly as representing an unchanged policy stance, in the sense of a consistent policy regime designed to achieve the Fed's inflation goal in the long run. In this way, the rule-based interest rate path would provide a benchmark around which discretionary decisions could be made.[16] In any specific situation, more or less expansionary policies than indicated by a rule could be adopted. During a recession the FOMC might want to lean toward a lower short-term interest rate than was called for by a rule. For example, if the growth rates of real and nominal GDP were to increase—and thus to signal an interest rate increase—the Fed might choose to override this signal if the *level* of real GDP were considered to be far below its potential level. By the same token, if the economy seemed to be "overheating," as a result, say, of a surge in demand for our exports, policy could lean in the direction of tightness for a time. So long as such discretionary deviations from a rule-based policy averaged out to zero over time, the long-run benefits of a feedback rule for inflation would be realized. Of course, if it were deemed advisable to change the inflation objective, the policy regime could be modified by changing the nominal GDP target itself.

In this paper we have focused on the general issue of whether using a nominal GDP feedback rule as a baseline for discretionary decisions might help the FOMC achieve its goals by rationalizing and simplifying the decisionmaking process. Of course, a number of practical issues would need to be addressed before such an approach could be adopted in practice. The biggest one would be to choose a specific nominal GDP feedback rule. As noted above, our earlier research suggests that a rule defined in terms of an interest rate instrument, a nominal GDP growth rate target, and a relatively mild reaction coefficient seems promising. However, since other researchers have supported other types of feedback rules, this issue is by no means

16. Following such an approach might enhance the credibility of the disinflation goal (Judd and Beebe, this issue.)

settled. Actual use of a feedback rule would require more research within the context of other models in order to narrow the range of appropriate choices of rules.

A number of more detailed issues also would arise. For example, the FOMC meets eight times per year, whereas the rules discussed above give a "recommendation" for the average level of the short-term interest rate over a quarterly period (given last quarter's level). Thus a method would need to be devised to link the decision period between FOMC meetings (which averages 6½ weeks) with the quarterly period of time used to define the rule.

Finally, as with any new approach to policy, there is no way to anticipate all of the problems that might be encountered if it actually were implemented. The process of implementation most likely would involve a good deal of learning and modification. The approach discussed in this paper does *not* require mechanically following a rule. Instead, it represents a discretionary approach that would be informed by a rule. As such, policymakers would continue to be in a position to use their judgment to react to circumstances as they arose, but with the benefit of the additional information provided by the rule.

REFERENCES

Barro, Robert J. 1986. "Recent Developments in the Theory of Rules Versus Discretion." *The Economic Journal* (Supplement) pp. 23-37.

Bureau of Economic Analysis, U.S. Department of Commerce. 1993. "Gross Domestic Product: Second Quarter 1993 (Advance)," News Release (July 29).

Gagnon, Joseph E., and Ralph W. Tryon. 1993. "Price and Output Stability Under Alternative Monetary Policy Rules." Board of Governors of the Federal Reserve System, *Finance and Economic Discussion Series,* Working Studies 1: Part 2.

Hallman, Jeffrey J., Richard D. Porter, and David Small. 1991. "Is the Price Level Tied to the M2 Monetary Aggregate in the Long Run?" *American Economic Review* (September) pp. 841-858.

Hess, Gregory D., David H. Small, and Flint Brayton. 1993. "Nominal Income Targeting with the Monetary Base as Instrument: An Evaluation of the McCallum Rule." Board of Governors of the Federal Reserve System, *Finance and Economics Discussion Series,* Working Studies 1: Part 2.

Howitt, Peter. 1990. "Zero Inflation as a Long-Term Target for Monetary Policy." In *Zero Inflation: The Goal of Price Stability* (Policy Study 8), ed. Richard G. Lipsey, pp. 67-108. Ottawa, Ontario: C.D. Howe Institute.

Judd, John P., and Jack H. Beebe. 1993. "The Output-Inflation Tradeoff in the United States: Has It Changed Since the Late 1970s?" Federal Reserve Bank of San Francisco *Economic Review,* this issue.

_____, and Brian Motley. 1991. "Nominal Feedback Rules for Monetary Policy," Federal Reserve Bank of San Francisco *Economic Review* (Summer) pp. 3-17.

_____, and _____. 1992. "Controlling Inflation with an Interest Rate Instrument." Federal Reserve Bank of San Francisco *Economic Review* 3, pp. 3-22.

_____, and Bharat Trehan. 1992. "Money, Credit and M2." Federal Reserve Bank of San Francisco *Weekly Letter* (September 4).

Kasman, Bruce. 1993. "A Comparison of Monetary Policy Operating Procedures in Six Industrial Countries." *Finance and Economics Discussion Series,* Working Studies 1: Part 1, March.

McCallum, Bennett T. 1990. "Targets, Indicators, and Instruments of Monetary Policy." *NBER Reprint* No. 1550.

Meltzer, Allan H. 1987. "The Limits of Short-Run Stabilization Policy." *Economic Inquiry* (January) pp. 1-14.

Motley, Brian. 1993. "Growth and Inflation: A Cross-Country Study." Working Paper 93-11. Federal Reserve Bank of San Francisco.

Taylor, John B. 1985. "What Would Nominal GDP Targeting Do to the Business Cycle?" *Carnegie-Rochester Series on Public Policy 22,* pp. 61-84.

_____. 1992. "Discretion Versus Policy Rules in Practice." CEPR Publication No. 327. Center for Economic Policy Research, Stanford University.

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

Money, Interest Rates and Economic Activity: Stylized Facts for Japan

Ramon Moreno and Sun Bae Kim

Economists, Federal Reserve Bank of San Francisco.

The authors thank the Editorial Committee, comprising Chan Huh, Elizabeth Laderman, and Bharat Trehan, for many helpful comments. We also thank Greg Holmes and Jacob Pozharny for research assistance.

This paper examines how financial market changes affect the usefulness of two alternative indicators of monetary policy in Japan, a monetary aggregate and an interest rate. The paper tests whether these variables are good predictors of output, and whether responses to shocks to these variables broadly conform to the implications of the monetary transmission model, over two periods between 1960 and 1992. In the earlier period when Japan's financial markets were less developed, a monetary aggregate (M2 + CDs) is a relatively useful indicator of monetary policy whereas an interest rate variable is not. In particular, we find some evidence of a "liquidity effect" in response to innovations in money. Neither variable is an entirely satisfactory indicator of monetary policy in the second sample. The results suggest that financial market development may have contributed to reducing the usefulness of money as an indicator of monetary policy.

Are monetary policy innovations better represented by shocks to money or to interest rates? In order to resolve this question, researchers have used vector autoregression (VAR) models to attempt to ascertain whether responses to innovations in either variable satisfy two criteria. First, changes or innovations in monetary policy should be good predictors of real economic activity. Second, the qualitative effects of monetary policy innovations should conform to those predicted by the traditional monetary transmission model; namely, in the short run, an expansionary policy leads to an excess supply of money because of output and price rigidities. In response to this excess supply, nominal and real interest rates fall. This "liquidity effect" is a key element in the monetary transmission mechanism which ultimately leads to an increase in real output.

The empirical evidence is ambiguous. Research applying VAR models to U.S. data generally concludes that interest rates are better predictors of real output than are monetary aggregates.[1] However, responses to shocks to either money or interest rates are not entirely consistent with the monetary transmission model. For example, some impulse response analyses reveal that while a positive monetary aggregate shock is associated with increases in the price level, it is also associated with an *increase* in interest rates, or no liquidity effect (Sims 1986, Leeper and Gordon 1992), and a *contraction* in output (Sims 1986). The last two responses do not correspond to the monetary transmission model.[2] Furthermore, while an interest rate innovation is associated with declines in money and output, as would be expected, it is also associated with an *increase* in the price level, a result that contradicts the monetary transmission model (Sims 1986).

---

1. See, for example, Sims (1980) and Bernanke and Blinder (1992). An exception is Strongin (1992) which focuses on a special representation of a very narrow monetary aggregate.

2. However, recent research indicates that a liquidity effect can be found in U.S. data if a narrower monetary aggregate is used that takes some of the subtleties of Fed operating procedure into account (Eichenbaum 1992, Christiano and Eichenbaum 1992, and Strongin 1992). While Eichenbaum (1992) finds a liquidity effect using a nonborrowed reserves aggregate, this measure is not entirely successful because real output declines in response to innovations in nonborrowed reserves. Strongin (1992) uses the ratio of nonborrowed reserves to total reserves as an indicator of policy innovations, and a recursive ordering that appears to successfully reflect Fed operating procedure. His indicator satisfies the two criteria outlined in the text.

Research applying VAR models to Japanese data also has yielded ambiguous results. For example, Suzuki, Kuroda, and Shirakawa (1988) find that a broad monetary aggregate, M2 + CDs, is a good predictor of Japanese real GNP.[3] However, the dynamic responses to money shocks presented in Sims (1992) do not conform to the transmission model. As in the U.S., the response to a money shock reveals the absence of the "liquidity" effect in Japan, as well as a contraction in output. Japanese data also yield the "price puzzle" of an interest rate innovation leading to a price increase.

Two explanations may be offered for why VAR models have failed to identify an unambiguous indicator of monetary policy in the U.S. or in Japan. First, the studies cited in this paper generally rely on a recursive identification procedure to distinguish between innovations in money or interest rates. Under certain conditions, innovations identified using such a procedure will not successfully distinguish monetary policy shocks from real shocks to aggregate supply or shocks to money demand. Difficulties in isolating shocks are particularly likely if monetary authorities do not consistently target a monetary aggregate or an interest rate. Second, developments in financial markets may influence the ability to identify a policy indicator. For example, even if the authorities consistently target a monetary aggregate, demand shocks may still cause short-term fluctuations in monetary aggregates if deregulation and innovations in financial markets weaken the central bank's effectiveness in controlling the monetary aggregate target. In fact, the choice of target itself may shift as a result of significant changes in financial markets.

Japan provides a potentially illuminating case study to ascertain the plausibility of the second explanation. Up to the early 1980s, securities markets were undeveloped and commercial banks were heavily dependent on the Bank of Japan (BOJ), which used both market and nonmarket instruments to achieve a credit target. Subsequent deregulation and innovations in financial markets have reduced the dependence of commercial banks on the BOJ. As a result, the BOJ has relied more heavily on market instruments for monetary control, and has paid attention to both monetary aggregates and interest rates.

The primary aim of this paper is to explore the implications of financial market changes for the identification of monetary policy innovations in Japan. We do this by estimating a four-variable VAR model of the Japanese economy similar to the models estimated by Sims (1980, 1992) over two sample periods: the first, 1960–1980, when securities markets were undeveloped and the BOJ wielded much greater direct influence on commercial banks in implementing monetary policy, and the second, 1981–1992, when Japan's financial markets became more developed, the dependence of banks on the BOJ declined, and the BOJ began to rely more heavily on market-based mechanisms for monetary control.

Our main findings may be summarized as follows. A monetary aggregate (M2 + CDs) is a relatively useful indicator of monetary policy in the first sample period. Money is a good predictor of output according to one measure used, and responses to money shocks also conform broadly to the implications of the monetary transmission model. In particular, we find some evidence of a "liquidity effect." In this sample, an interest rate indicator has about the same predictive power as money, but responses to interest rate shocks do not entirely conform to the monetary transmission model.

In contrast to the first sample, no entirely successful indicator of monetary policy is identified in the second sample. Although money continues to be a good predictor of output in this sample (much better than interest rates), the responses to innovations in either variable cannot be interpreted as reflecting innovations in monetary policy.

The rest of the paper is organized as follows. Section I describes how financial markets and the approach to monetary policy have changed over time in Japan. Section II describes the macroeconomic models to be estimated, discusses the approach to identification, motivates the selection of variables included in the two alternative models, and summarizes the estimation procedure. Section III reports the results while Section IV provides some conclusions.

I. JAPANESE MONETARY POLICY: INSTRUMENTS AND TARGETS

The monetary transmission model implies a certain relationship among variables that can be influenced by policy—monetary aggregates and interest rates—and the ultimate objectives of policy, such as real economic activity. One issue confronting policymakers is whether to target a monetary aggregate or an interest rate.[4] This policy choice determines whether, in empirical analysis, a monetary aggregate or an interest rate will be a suitable measure of changes in monetary policy. An interest rate (e.g., the

3. The authors base this conclusion on exclusion restrictions or "Granger causality." Ito (1982) finds that the narrower monetary aggregate M1 is not a good predictor of Japanese output in a VAR model, according to variance decompositions. Other studies focus on the ability of money to predict nominal GNP. See the survey by Okina (1985).

4. For a survey of this issue see Friedman (1990).

interbank rate) may be considered a useful indicator of monetary policy if policymakers supply reserves perfectly elastically to target a given interest rate. In this case, shocks to money demand will not affect the targeted rate. A monetary aggregate will be the appropriate indicator of monetary policy if monetary authorities target the aggregate and do not accommodate shocks to demand.[5] Thus the use of either an interest rate or a monetary aggregate as an indicator of monetary policy involves fairly stringent assumptions about the behavior of monetary authorities. In practice, the targets and instruments used by monetary authorities, and the commitment to any given target, vary over time. Financial market development may also affect the ability to control monetary aggregates with precision.

These various considerations make empirical analysis difficult. However, such difficulties may be mitigated by examining the institutional setting and the operating procedure employed by the BOJ, as well as the financial market environment. Such an examination may facilitate the choice of an indicator of monetary policy and the interpretation of any empirical results.[6] In this section we perform such an institutional review, focusing on the following questions. First, has monetary policy in Japan primarily targeted a monetary aggregate or interest rate(s)? Second, given a choice of a particular target (or set of targets), what operational procedure has been used to implement it? Third, how has financial liberalization affected monetary control and the choice of operational procedure and target? To begin, we provide some background on the postwar Japanese financial system.

Japan's Financial System

Throughout most of the postwar period, the task of mobilizing funds from net savers to investors in Japan has fallen primarily on the banking sector. Up to the 1980s, the government intervened actively in this process of intermediation, initially in an effort to promote investment and growth, and later to meet the financing needs of the government. Deposit rates were strictly controlled, while the bank lending rate was anchored to the official discount rate, ensuring a relatively tight link between government interest rate policy and the cost of funds faced by nonbank borrowers. Selective and aggregate credit targeting ensured direct government input in the allocation and growth in credit, as did the heavy reliance of banks on borrowing from the BOJ's discount window. Controls over the nonbank financial sector ensured that market forces would not erode the effectiveness of banking sector controls. For example, the government controlled which firms could issue bonds as well as the corporate bond rate. Finally, exchange and capital controls prevented financial market participants from circumventing regulation via overseas transactions.

Japan's regulated financial system gradually gave way to market forces beginning in the mid-1970s. The main factor appears to have been the slowdown in economic growth in the early 1970s, which sharply reduced government revenues and prompted large increases in government borrowing. Initially, government bonds were allocated to banks at below-market interest rates, but as the volume of borrowing increased, there was strong pressure for the development of an active secondary market in government bonds at a market-determined rate. By arbitrage, this stimulated the development of short-term money markets such as the *gensaki* (repurchase) and CD markets. However, a short-term government debt market did not develop until very recently. The liberalization of financial markets extended to international capital transactions, restrictions on which were progressively dismantled in the course of the 1980s.[7]

Monetary Control Prior to Deregulation

In an environment in which banks dominated as financial intermediaries and the heavily indebted private business sector had virtually no alternative to bank loans for external financing, the BOJ's traditional approach to monetary control consisted of controlling the amount of bank lending to the nonfinancial corporate sector. We can think of this as similar to monetary targeting because, with corporate loans dominating the asset side of banks' balance sheets, bank credit and broad monetary aggregates tended to move very closely for much of the postwar period. While it is clear from accounts of BOJ practices that the BOJ attached a great deal of importance to achieving its credit targets, these credit targets have not been disclosed, so it is unclear whether they were truly exogenous, or whether the BOJ from time to time accommodated shocks to credit demand. Also, the weight the BOJ attached to curbing inflation

5. There may be ambiguities in interpreting the stance of monetary policy when using an interest rate target. Since shocks to demand are fully accommodated, an unchanged interest rate may be consistent with stimulus or contraction in output. Similar difficulties in interpretation may arise when focusing on a money aggregate if money demand is unstable. Such instability became a concern in Japan in the second half of the 1980s, when money growth accelerated with less-than-proportionate increases in nominal income.

6. For example, Strongin (1992) motivates a recursive identification procedure from knowledge of the Federal Reserve's operational procedures. However, his reasoning is not applicable to Japan.

7. For overviews of the postwar Japanese financial system and the process of deregulation, see Feldman (1986) and Hamada and Horiuchi (1987).

prior to the mid-1970s is uncertain, whereas it is apparent that a high weight was attached to curbing inflation since then.

To achieve its credit objectives, the BOJ relied largely on two instruments. First, it sought to influence interest rates in the interbank market. Second, it provided direct guidelines for commercial bank lending.[8]

Interbank Interest Rates. On a day-to-day basis, the BOJ sought to influence the supply of credit and money by targeting the call money rate in the interbank market.[9] To a large extent, the transmission mechanism relied upon interest rate rigidities in the system and the heavy dependence of the corporate sector on bank lending. Loan as well as deposit rates were subject to administrative controls, so banks could not easily pass on to corporate borrowers changes in the interbank rates. In this setting, changes in the call money rate had a direct and immediate impact on bank profitability, and consequently on the growth of money and credit. For example, a rise in the call rate resulting from BOJ tightening would reduce the marginal profitability of lending. In response, banks would ration credit, forcing corporations to curtail investment, and the process would ultimately result in a reduction in the broad money supply.[10]

Accounts of BOJ's operating procedures suggest that reserves were supplied elastically at the call market rate consistent with the targeted level of credit. Monetary authorities paid particular attention to the "reserve progress ratio," which measures reserves accumulated by banks relative to those required within a maintenance period.[11] Notably, call transactions in Japan involved money market brokers (*Tanshi kaisha*). These brokers maintain close informational contact with the BOJ and, in close consultation with the BOJ, set the rate at the opening of the markets each day. If the initially quoted rate failed to equate demand and supply, the BOJ typically would adjust the supply of reserves to achieve equilibrium at its target interest rate.[12]

In the absence of a short-term market for government debt, discount window lending by the BOJ was the main instrument for short-run adjustments of bank reserves. A rationing scheme governed this method of monetary control. The BOJ provided loans to financial institutions (mainly city banks) at the official discount rate (ODR), typically at a rate below the interest rates in the interbank market. BOJ lending thus amounted to a subsidy and Japanese banks naturally preferred to rely on the central bank for liquidity.[13] Consequently, in contrast to the U.S., where the ratio of borrowed reserves to required reserves seldom exceeds 5 percent, the level of discount window borrowing by Japanese banks often has exceeded the level of required reserves.

Direct Control of Bank Credit. Another instrument of monetary control by the BOJ was the direct quantitative control of commercial bank lending through so-called "window guidance." To tighten the supply of money and credit, the BOJ would impose individual ceilings on new lending by commercial banks, in particular, the city banks. In formulating these ceilings, the authorities used information garnered during day-to-day contacts through deposit and lending transactions with individual financial institutions, such as their future loan plans and prospective fund positions. In addition, the BOJ received from city banks reports on a longer-term basis (monthly until 1963 and quarterly thereafter) which included forecasts of future fund-raising activity and the outlook for deposits and loans.

The BOJ had a number of ways to dissuade banks from lending in excess of their prescribed ceiling, such as curbing its discount window lending, thus compelling a bank to borrow in the more expensive call money market or to sell commercial bills. In practice, banks complied with BOJ guidelines with little need of persuasion because of their heavy reliance on BOJ discount window loans for their funds. Thus, according to Suzuki (1980), in no case did a bank exceed the limits imposed up to the late 1970s.

Monetary Control under Financial Liberalization

The mid-1970s initiated a process of financial deregulation and innovation that continues to date. From the vantage point of conducting monetary policy, three changes have been particularly significant. First, the importance of bank

8. The BOJ occasionally also resorted to changes in required reserves. These changes were relatively infrequent, and the use of this instrument was discontinued in 1982.

9. The call market rate is a short-term market comparable to the Federal funds market in the U.S. It is still the BOJ's primary operating target. There is also a bill discount market where commercial bills are discounted.

10. For an authoritative discussion of the transmission channels of monetary policy in Japan, see Suzuki (1980), parts II, III, and IV.

11. The reserve maintenance period in Japan is one month that straddles two calendar months. It runs from the 16th day of a month through the 15th day of the following month.

12. Under such a system, interbank rates could not fluctuate on a daily basis. See Dotsey (1986) and Fukui (1986) for details. Similar operating arrangements are in place today except that interbank rates fluctuate more freely than in the past. See footnote 18.

13. The BOJ decided on the level of bank borrowing (up to a predetermined quarterly ceiling), the term of borrowing, and the interest rate associated with borrowing. Officials at the BOJ could call up each city bank as frequently as daily to indicate how much they could borrow.

loans has sharply declined as a result of the slowdown in corporate investment and the move toward securitization. This, coupled with the large flotation of government bonds, has weakened the link between corporate lending and the monetary aggregate. Second, banks have been able to raise funds from a wider array of financial instruments and markets, such as the CD and euroyen markets, thus reducing their reliance on BOJ credit and eroding BOJ leverage in using credit rationing under window guidance. Third, assets with market-determined prices have come to predominate the portfolios of all sectors of the economy. The disintermediation between administered and market-priced assets has weakened BOJ's traditional transmission channel of altering the spread between interbank rates, on the one hand, and the administered loan and deposit rates, on the other.

These developments have led to a gradual shift from the late 1970s through the 1980s in the objectives of monetary policy. The BOJ began to pay more direct attention to the behavior of monetary—as opposed to credit—aggregates, and in 1978 began announcing "forecasts" of the growth in M2 + CDs. However, it is not clear that the BOJ has fully embraced monetary targeting, as might be inferred from the writings of some influential observers. (Friedman 1985, Meltzer 1986.) First, as BOJ officials emphasize repeatedly, these announced figures are projections rather than targets.[14] There is evidence suggesting that the BOJ has not tried systematically to offset differences between actual and targeted money (Ito 1989, Judd and Hutchison 1992). Second, broad money has not been the sole target, but has served as one of the primary indicators among a group of financial variables (Hamada and Hayashi 1985, Hutchison 1986, and Kasman and Rodrigues 1991). In fact, the BOJ appears to have gradually reduced its emphasis on broad money in recent years while focusing a great deal of attention on market interest rates (e.g. the call rate, the *gensaki* or repurchase rate, and the CD rate).[15]

Concomitant with these changes in the policy targets or indicators, the BOJ's operational procedures have evolved as well. The use of discount window guidance, in the form of the central bank instructing individual banks in their lending plans, was curtailed, although a more limited form of window guidance, through which the BOJ communicated its aggregate lending plans and overall policy stance to individual banks, continued into 1991.

In the wake of financial liberalization, the BOJ also extended its intervention outside the interbank market. The BOJ began open market operations in CDs in 1986, *gensaki* in 1987, and commercial paper in 1989, when active operations in short-term government securities were also initiated.[16] Nevertheless, the call money rate (along with the interbank commercial bill rate) is still the most important interest rate target of the BOJ, and BOJ discount window lending is still the primary mechanism for regulating the quantity of bank reserves.[17] Indeed, the effectiveness of the call market appears to have been enhanced by the greater flexibility in interbank interest rates resulting from deregulation in the late 1970s.[18]

Monetary Control: Summary and Implications

Our discussion sheds light on two important elements of the BOJ's approach to monetary policy in the postwar period. First, with respect to targeting, there is some uncertainty about the precise features of BOJ credit or money targeting. During the earlier periods, credit targets were reportedly consistently met. However, these targets were not announced, so it is unclear whether they were exogenous or adjusted to accommodate shocks to money demand. There also has been disagreement on whether the Bank of Japan actually adopted monetary targeting after 1978. However, BOJ statements suggest that there was no strict targeting as such, and that the behavior of monetary

14. For example, Suzuki states (1986, pp.192-193) that the BOJ focuses on control of the broad money supply but does not set a target. He indicates further that the BOJ "does not follow an operating procedure rule of constant money supply growth." Also, as noted by Judd and Hutchison (1992), these "annual" forecasts which were announced on a quarterly basis actually extended only one quarter ahead; that is, the forecasted annual growth rates are averages of three quarters of realized, and one quarter of projected, money growth. As such, these projections contained relatively little new information on monetary policy stance as compared to, say, the Federal Reserve's targets.

15. The increased attention paid to market interest rates is officially acknowledged by the BOJ staff itself. Nakao and Horii (1991) note, for instance, that the BOJ "has increased the number of reservations it applies to its interpretation of money supply, bank credit and other volume indicators of finance. . . [while] it has enhanced attention given to market interest rate developments."

16. Because of the relatively underdeveloped market for short-term government securities, open market operations in Japan necessarily have relied upon private short-term instruments.

17. See for example, Ohkubo (1983), Suzuki (1986), and Suzuki, et al. (1988).

18. Notable changes in the interbank market include: allowance for more frequent quotations on the call rate and the resale of bills allowed after one month from purchase (June 1978); the introduction of seven-day call money with a freely determined interest rate (October 1978); and the introduction of one-month bills at unregulated rates (November 1978). The process of liberalizing the interbank market was largely concluded in 1979 with the abolition of quotation systems in the call market and the introduction of shorter-term (2-6 days) call money (April) and the liberalization of rates on 2-month bills (October). From late 1979, therefore, rates in both the call and bill market have fluctuated daily (Dotsey 1986).

aggregates was monitored along with other indicators, such as interest rates. BOJ statements also indicate that in recent years the weight assigned to interest rates has increased.

Second, in the credit-dominated regime prior to the onset of financial deregulation and innovation, the BOJ depended more heavily on nonmarket mechanisms for monetary control, such as window guidance, to control money and credit. There were also few substitutes for bank deposits, so portfolio shifts were less likely to affect the behavior of monetary aggregates. It seems plausible to argue that the BOJ's ability to control monetary aggregates precisely was greater during this earlier period than it was later, when the BOJ deemphasized nonmarket mechanisms for control and when the development of financial markets broadened the spectrum of assets available to savers.

Given these characteristics, it seems reasonable to expect that monetary aggregates are unlikely to serve as a good indicator of monetary policy since the late 1970s or early 1980s, when financial innovation began in Japan. Monetary aggregates will serve as a good indicator in the earlier period if credit targets did not accommodate demand shocks which, as stated previously, is not entirely clear. The reason is that during this earlier period financial markets were relatively less developed and the BOJ relied more heavily on direct, nonmarket instruments that are likely to have significantly enhanced the precision of its control. As for interest rates, our institutional review provides no clearcut basis for deciding whether they might serve as useful policy indicators. However, the reader may note that the BOJ appears to have consistently used an interest rate as an operating target, and appears to have paid closer attention to the implications of interest rates for aggregate economic activity as the 1980s progressed. The empirical analysis may clarify some of these uncertainties.

II. THE MODEL

Structural Model and Identification

To motivate the approach followed in this paper, consider an economy described by a vector of nonpolicy variables, Z_t, that may be represented by indicators of economic activity like output and inflation, and a vector of policy variables, I_t, that can be influenced by monetary authorities, such as a monetary aggregate and an interest rate. The interaction by the variables is summarized by the following two equations[19]

$$(1) \quad Z_t = B_0 Z_t + B_1 Z_{t-1} + C_0 I_t + C_1 I_{t-1} + u_t$$

$$(2) \quad I_t = D_0 Z_t + D_1 Z_{t-1} + G I_{t-1} + v_t$$

where u, v are orthogonal disturbances.

One way of identifying this model is to assume that contemporaneous I does not enter equation (1) ($C_0 = 0$), so policy actions affect real variables only with a lag. However, policymakers respond to contemporaneous innovations in macroeconomic activity.

$$(3) \quad Z_t = (I - B_0)^{-1} [B_1 Z_{t-1} + C_1 I_{t-1} + u_t]$$

$$(4) \quad I_t = (D_1 + D_0(I - B_0)^{-1} B_1) Z_{t-1} + (G + D_0(I - B_0)^{-1} C_1) I_{t-1} + v_t + D_0(I - B_0)^{-1} u_t$$

In equation (4) I_t is contemporaneously affected by the policy innovation v_t and also by contemporaneous macroeconomic shocks u_t. Identification as proposed above can be obtained by estimating a VAR comprising Z, I. The orthogonalized innovations that satisfy the recursive structure assumed in equations (3) and (4) can then be identified by applying the Choleski decomposition to the variance-covariance matrix of the residuals, putting I last in the ordering.

The model described above can be used to assess the predictive ability of alternative possible measures of monetary policy by: (i) testing for the significance of exclusion restrictions on monetary policy variables in the industrial production equation, (ii) estimating the respective contributions of orthogonalized innovations in money or interest rates to the variance of the forecast error of output at various horizons. Similar procedures for assessing predictive ability are used by Sims (1980, 1992) and by Bernanke and Blinder (1992).

To ascertain whether responses to monetary variables conform to theoretical expectations, the VAR model can be inverted to obtain the impulse responses (the coefficients of the moving average representation of the model) to orthogonalized innovations in money and interest rates. A comparison of these responses can then be used to assess the extent to which the responses to innovations in these variables conform to the monetary transmission model described earlier.

As is well known, the Choleski identification procedure proposed here has been criticized on a number of grounds. One potential difficulty is that the results can be sensitive to the ordering of the variables.[20] Another difficulty, reflected in the sometimes counterintuitive responses to

19. See Bernanke and Blinder (1992) for a similar discussion.

20. This criticism does not apply to the model used in this paper because the contemporaneous correlations of the residuals of the estimated VAR model are low.

shocks cited earlier in the introduction, is that it is not entirely clear how innovations identified using this procedure are to be interpreted. For example, orthogonalized innovations in money may reflect shocks to money supply or money demand.

At least two responses may be offered to these criticisms. First, the application of these identification methods to Japanese data can actually shed further light on the plausibility of the Choleski identification procedure, which is still widely used in empirical VAR models of the U.S. (The users include authors who have also used alternative identification procedures explicitly based on economic theory, such as Bernanke and Sims). If the responses to shocks broadly conform to theoretical expectations, it can be argued that the Choleski procedure is a reasonable approximation to a model that is identified on the basis of economic theory.

Second, efforts to identify VAR models using economic theory have also been criticized on various grounds. As pointed out by Sims (1986), any empirical study raises debatable questions about identification that will leave readers more or less uncomfortable about applying the conclusions. Under these circumstances, researchers may be well advised to experiment with different approaches to identification. The present paper can then be seen as one step in allowing the data "to speak" about Japanese macroeconomic behavior.[21] Future studies that attempt to utilize alternative identification procedures are not ruled out.

The Model

A four-variable monthly VAR model for Japan was estimated, comprising industrial production (IP) to represent output, the consumer price index (CPI) to represent price, a broad monetary aggregate (M2+CDs) to represent money, and the interbank call money rate (CMR) to represent interest rates. The variables were entered in this order in identifying the orthogonalized innovations.[22] The data span the period 1961.1–1992.8. Data and sources are described in the Appendix.

As discussed earlier, financial liberalization and changes in the Bank of Japan's approach to monetary policy are likely to have affected the relationship between money and interest rates and economic activity. This suggests that it would be desirable to estimate the model over two subsamples. The first subsample would correspond to the period when financial markets were undeveloped, commercial banks were heavily dependent on the BOJ for funding, and the BOJ emphasized credit targets and sought to influence bank behavior directly through window guidance. The second would correspond to the period when financial markets were more developed, commercial banks had more access to sources of financing outside the BOJ (including external financing), and the BOJ paid more attention to money and interest rates (rather than credit) and began to rely more heavily on market-based mechanisms for monetary control.

Given the gradual pace of financial innovation in Japan, there is no obvious single candidate for a break date. Likely candidates for a break date are somewhere between 1975, when large quantities of government bonds were first issued, and 1981, when Japan's foreign exchange controls were first liberalized. The date 1981.1, which is the first month after foreign exchange controls were liberalized in Japan, is selected as a reasonable candidate. At that time, the liberalization of controls loosened the dependence of Japanese commercial banks on the Bank of Japan by allowing them to draw on foreign sources of funding. Also, the impact of gradual financial liberalization on macroeconomic relationships is more likely to have been manifest by then. Thus, the first sample period spans 1961.1–1980.12, and the second, 1981.1–1992.8.

To account for non-stationarity in the data, the model was estimated in first differences of the logs of the variables, with the exception of the interest rate, where the first difference of the series was used.[23] As the frequency of the data was monthly, lag lengths were set at 12 for both subsamples. No other criterion was used in setting the lag length.

III. RESULTS

Predictive Ability

To assess the predictive ability of alternative monetary indicators, Table 1 reports the results of tests for exclusion restrictions on the right-hand-side variables of the output

21. Structural VAR models of the Japanese economy have been estimated by Hutchison and Walsh (1992), Hutchison (forthcoming), and Moreno (1992). However, none of these models are explicitly designed to analyze monetary policy.

22. The ordering places M2+CDs prior to CMR, which assumes that the former is contemporaneously unaffected by innovations to the latter. However, the correlation between residuals in the M2+CDs and the call money rate equations is small, about 8 percent in the first subsample and 18 percent in the second. Thus, the results are not likely to be very sensitive to the ordering assumed.

23. As is by now well-known, however, there is a great deal of uncertainty surrounding the stochastic properties of macroeconomic data (Cochrane 1991, Rudebusch 1992).

(industrial production) equation and the results of variance decompositions of the forecast error of output.

In the first subsample, tests of exclusion restrictions indicate that money is a good predictor of output; in fact, it is better than the interest rate. The evidence offered by variance decompositions is mixed. Money and the interest rate make similar (and relatively small) contributions to the variance of the forecast error after two years, after which the contribution of money falls off, while the contribution of the call money rate remains about the same. It is worth noting the Japanese economy experienced relatively large supply shocks during this period (notably, large declines in productivity and growth), and also fiscal policy shocks, and these may have tended to reduce the observed contribution of the monetary indicators (money or interest rates) to the variance of the forecast error of output.

For this sample, money and the interest rate are both generally good predictors of output, so the criterion of predictive ability does not allow us to choose unambiguously between the two. An analysis of responses to innovations in each of these variables is needed to shed further light on which variable may be a better indicator of monetary policy in Japan in this sample period. This is in contrast to Sims's (1980) well-known result for the U.S. case, which found that interest rates rob the monetary aggregate of predictive power. According to the criterion of predictive ability, in the U.S. the interest rate is favored over money as an indicator of monetary policy innovations.

In the second subsample, tests of exclusion restrictions as well as the variance decomposition results indicate that the monetary aggregate is a much better predictor of output than is the call money rate. At a two-year horizon, money accounts for 58 percent of the variance of the forecast error of output in the second sub-sample, compared to 1 percent for the call money rate. In fact, the predictive power of money is much larger in the second subsample (58 percent) than in the first subsample (13 percent).

Our findings on predictive ability are consistent with some previous studies of the Japanese economy that use similar techniques and also replicate some ambiguities observed in this literature. In particular, the finding that money is a good predictor of real GNP according to exclusion restrictions in both subsamples is consistent with Suzuki, et al. (1988) which findings are based on a five-variable VAR over the period 1968.Q1–1987.Q4.[24] It is also interesting that our finding that money is not a good predictor of output according to variance decompositions for the first subsample is similar to Ito's (1982) findings using M1 over a similar period.

TABLE 1

INDUSTRIAL PRODUCTION

	TESTS OF EXCLUSION RESTRICTIONS			
	IP	CPI	M2+CDs	CMR
1960.1-1980.12	8.7×10^{-6} **	.04 *	8.4×10^{-4} **	2.7×10^{-3} **
1981.1-1992.8	.02 *	.14	.04 *	.56

	VARIANCE DECOMPOSITION (PERCENT)				
	Months	IP	CPI	M2+CDs	CMR
1960.1-1980.12	12	79	8	11	1
	24	27	44	13	15
	36	16	60	8	16
1981.1-1992.8	12	54	7	39	1
	24	26	15	58	1
	36	15	21	64	0

NOTES:
** Significant at 1%
* Significant at 5%

To sum up, the preceding findings provide mixed evidence that both money and interest rates are good predictors of output in the first sample. Money is a much better predictor of output in the second sample. These findings support the view that money may be a useful indicator of monetary policy, perhaps better than the interest rate. However, such a conclusion is valid only if innovations to money satisfy the predictions of the monetary transmission model, while innovations in interest rates do not. We turn to this question now.

Do Responses Conform to the Monetary Model?

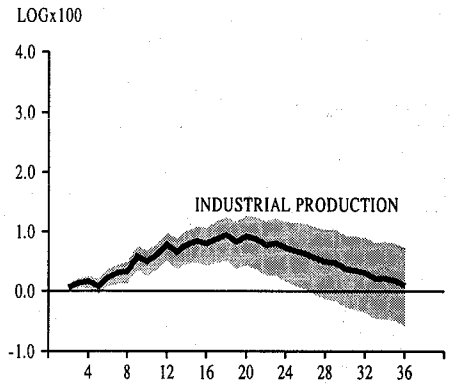
To assess whether innovations to money or interest rates conform to the predictions of the monetary transmission model, Figure 1 reports the point estimates of the responses of the variables to innovations in monetary aggregates in the first and second samples, while Figure 2 illustrates the corresponding responses to interest rate innovations. The shaded area illustrates one standard error band around

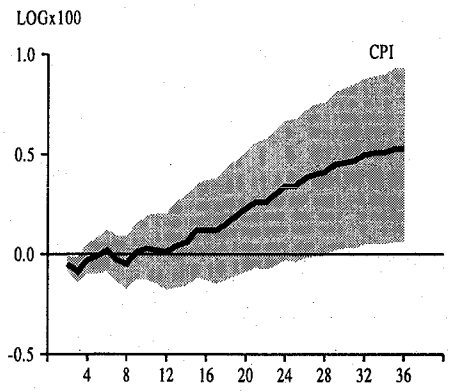
24. The Suzuki, et al. (1988) model comprises base money, the weighted average of the call money rate and the bill rate, M2 + CDs, real output, and the GNP deflator. Other studies have focused largely on the ability of money to predict nominal output. See also the survey by Okina (1985).

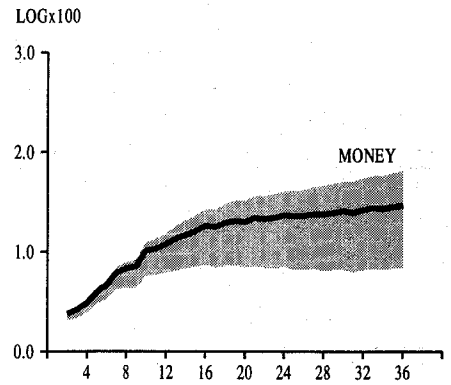
FIGURE 1

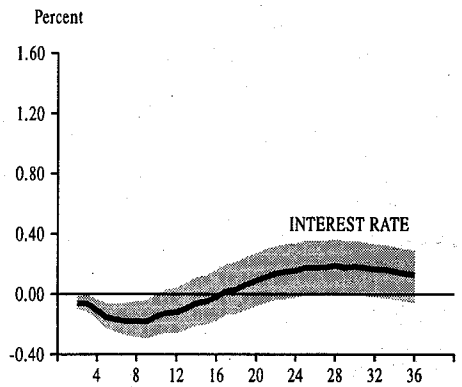
Responses to Innovations in Money

1960.1–1980.1

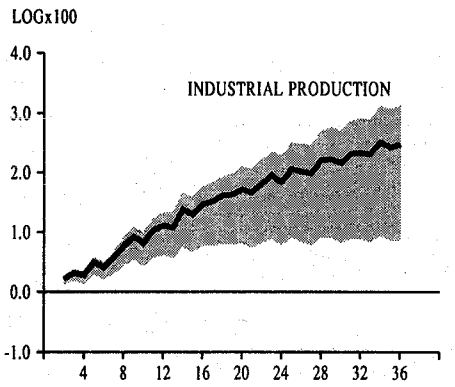


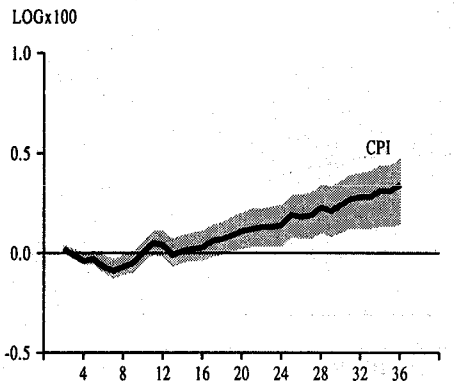


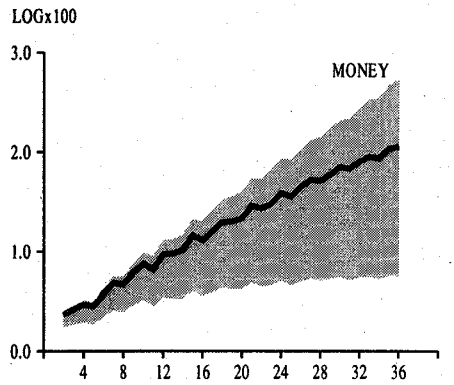




1981.1–1992.8







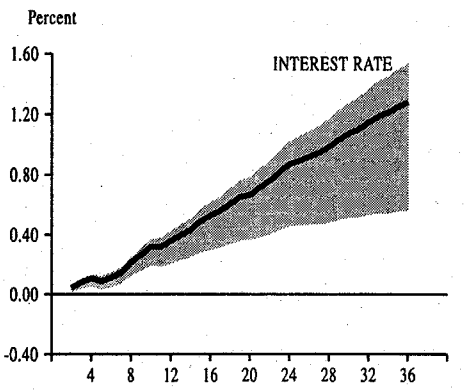
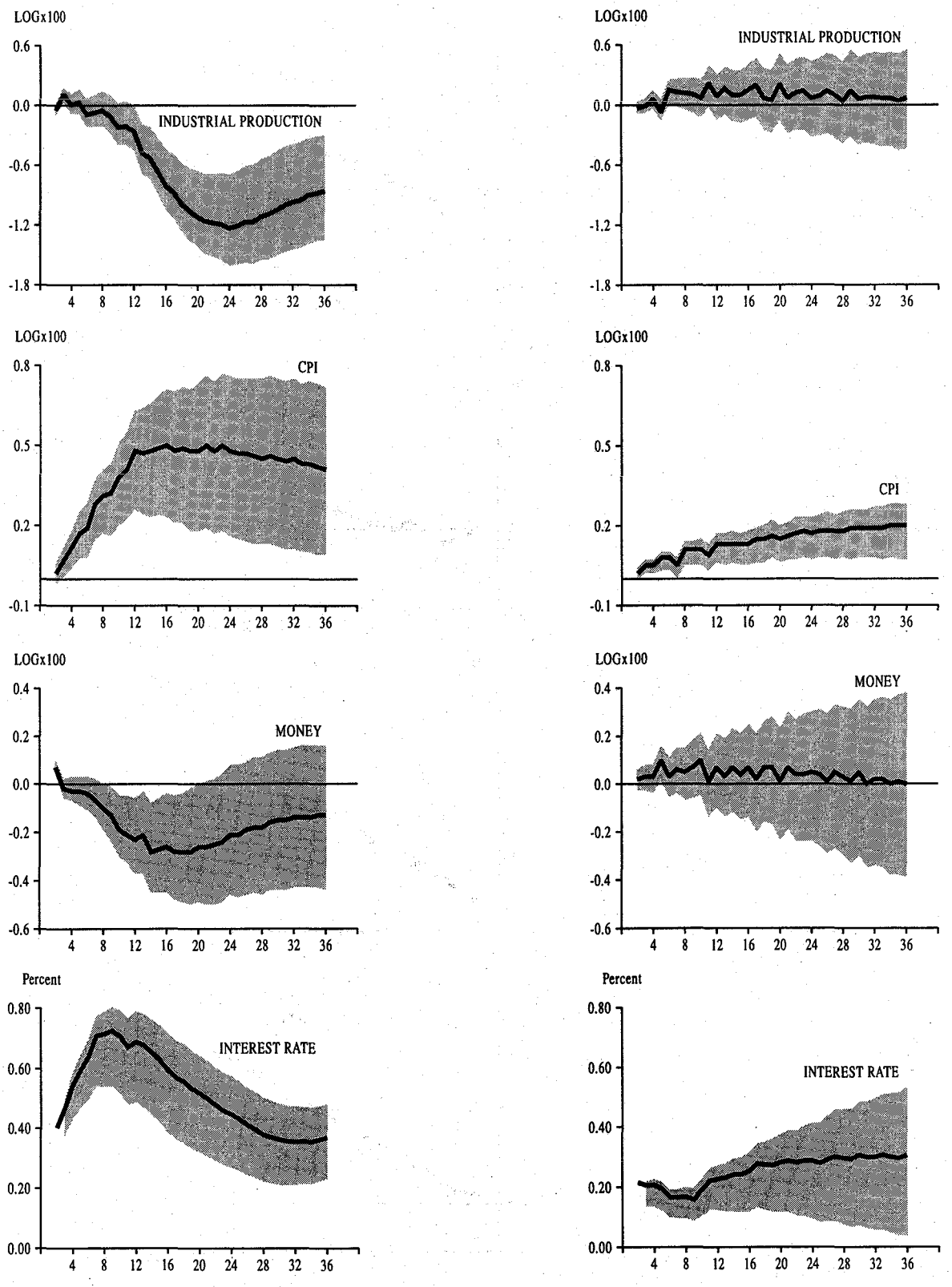


FIGURE 2

RESPONSES TO INNOVATIONS IN INTEREST RATES

1960.1–1980.1 1981.1–1992.8



the point estimate. The standard errors are obtained using a bootstrapping procedure.[25]

Focusing on the point estimates, in the first subsample an innovation to money is associated with a fall in the interest rate below its initial level that persists for about a year and a half, followed by a persistent rise that is subsequently reversed. Thus, there appears to be a "liquidity effect" in this model. Other responses to innovations in money are broadly consistent with the monetary transmission model. In response to a monetary aggregate shock, output rises temporarily, while prices rise after about one year.

In response to a positive interest rate innovation, money falls, and output falls. However, there is an extended increase in the price level that is difficult to interpret. A similar "price puzzle" was found by Sims (1992, 1986) for Japan and the U.S. While the responses to innovations in interest rates are difficult to interpret, one possibility is that they reflect aggregate supply shocks. This would account for a set of events where interest rates rise, output falls, and the price level rises. An alternative explanation, offered by Sims, is that interest rate innovations reflect an effort by monetary authorities to offset anticipated increases in prices. However, the inflationary pressure is not entirely offset so an increase in price is observed following the increase in the interest rate.

Figure 1 indicates that in contrast to the first subsample, in the second sub-sample an innovation in money is associated with an increase in the interest rate. Thus, there appears to be no liquidity effect. As in the first subsample, output, prices and money increase; however, the responses of output and money are much larger and more persistent than in the first subsample. The direction of the responses suggests that money innovations are demand shocks. However, the persistence of the responses suggests the presence of permanent supply shocks. Thus, innovations to money are not easily interpreted in this sample. Figure 2 reveals that all the variables in the model now rise in response to an innovation in interest rates (in the case of output, after an initial decline). These results suggest that an increase in interest rates reflects increases in the demand for money rather than a tightening in policy.

To sum up, the responses to innovations in money in the first sample fit the predictions of the monetary transmission model, but do not do so in the second sample. Responses to innovations in interest rates do not consistently fit the predictions of the monetary transmission model in either period. Taken together with the results on predictive ability cited earlier, money appears to be a relatively useful indicator of monetary policy shocks in the first sample. In contrast, the methods used in this paper do not successfully identify a monetary policy indicator in the second sample. The results for the first sample should be interpreted with caution, because estimates are in some cases imprecise. Nevertheless, they are quite interesting, as they suggest that financial market development may have contributed to reducing the usefulness of money as an indicator of monetary policy in Japan.

The preceding results may also be compared to those reported in Sims (1992) which studied several industrial economies, including Japan. Sims finds that the response of output to a monetary innovation is negative and is not associated with a liquidity effect. To isolate the reasons for the differences, we reestimated the model using M1 over two subsamples, and found that Sims's results differ from ours largely because he uses M1 rather than M2 + CDs as a monetary aggregate.

IV. CONCLUSIONS

This paper provides suggestive evidence that a broad monetary aggregate is a better indicator of monetary policy than an interest rate during the period when Japan's financial markets were less developed, the BOJ focused on a credit target and relied more heavily on nonmarket instruments for monetary control to achieve its policy target. During this period, there is mixed evidence that both a monetary aggregate and an interest rate are good predictors of output. However, analysis of the responses to shocks suggests that the monetary aggregate is the better indicator of monetary policy. Responses to money shocks broadly conform to the implications of the monetary transmission model, whereas the responses to interest rates do not. In particular, point estimates indicate the existence of a liquidity effect in response to innovations in a monetary aggregate. Responses to interest rate innovations suggest that such innovations may reflect aggregate supply shocks or policy responses to anticipated inflation.

In the second period, when Japan's financial markets became more developed, and the BOJ appeared to adopt a more eclectic approach to monetary policy, money is

25. To construct the standard error band, we bootstrapped the residuals of the VAR. The residuals were used to construct artificial series for the variables in the models. The VARs were then rerun using the artificial series and the impulse responses were recomputed. The simulations were repeated 1200 times. The one standard error band was computed by taking the square root of the mean squared deviation of the artificial impulse responses (above and below) from the point estimate at each step. By construction, the impulse responses obtained using the original data are inside the band. They are also asymmetric. This resembles the procedure used by Blanchard and Quah (1989). It may be noted that in a few cases, the actual responses to shocks fell partly outside the space spanned by the artificial impulse responses. In these cases, the standard error bands are not shaded over the applicable horizons. This difficulty should be borne in mind in interpreting the results.

unambiguously a good predictor of output, whereas the interest rate is not. However this does not imply that money is a good indicator of policy, as the responses to innovations do not conform to the monetary transmission model. In particular, there is no evidence of a liquidity effect, suggesting that money innovations are better interpreted as reflecting shocks to demand rather than policy changes. Interest rate innovations also appear to reflect shocks to demand.

As is often the case in this type of analysis, the estimates are in some cases imprecise and should be interpreted with caution. Nevertheless, they are quite interesting, as they suggest that financial market development may have contributed to reducing the usefulness of money as an indicator of monetary policy. As discussed in our institutional review, this may have occurred because financial market development encouraged the Bank of Japan to pay greater attention to interest rates and also loosened the control of monetary authorities over the monetary aggregate. Similar forces may explain why researchers have had difficulty in identifying an indicator of monetary policy in the U.S., where financial markets have been much more developed than Japanese financial markets in the postwar period.

APPENDIX

Data Description and Sources

Consumer Price Index, Seasonally Adjusted (CPI): Index of consumer prices covering the whole country excluding single-person households and those engaged mainly in agriculture, forestry, and fishing. Base year is 1985. Seasonally adjusted by Federal Reserve Bank of San Francisco staff using X-11 filter. Source: International Monetary Fund, *International Financial Statistics* (IFS).

Industrial Production, Seasonally Adjusted (IP): Index of monthly production by 9 mining and 523 manufacturing industries, weighted by 1985 value-added data. Base year is 1985. Seasonally adjusted by Federal Reserve Bank of San Francisco staff using X-11 filter. Source: IFS.

Call Money Rate (CMR). Rate in interbank call money market. Source: Bank of Japan.

M2 + CDs, Seasonally Adjusted: M1 plus quasi-money (time, savings, and foreign currency deposits of resident sectors other than central government) plus certificates of deposit in trillions of yen. End of month. Seasonally adjusted by Federal Reserve Bank of San Francisco Staff by applying X-11 filter. Source: IFS.

REFERENCES

Bernanke, Ben, and Alan Blinder. 1992. "The Federal Funds Rate and the Channels of Monetary Transmission." *American Economic Review* 82, pp. 901-921.

Blanchard, Olivier Jean, and Danny Quah. 1989. "The Dynamic Effects of Aggregate Demand and Supply Disturbances." *American Economic Review* 79, pp. 655-673.

Bryant, R. 1991. "Model Representations of Japanese Monetary Policy." *Bank of Japan Monetary and Economic Studies* 9, pp. 11-61.

Christiano, Lawrence J., and Martin Eichenbaum. 1992. "Liquidity Effects, Monetary Policy and the Business Cycle." NBER Working Paper No. 4129.

Cochrane, John. 1991. "A Critique of the Application of Unit Root Tests." *Journal of Economic Dynamics and Control* 15, pp. 275-284.

Dotsey, Michael. 1986. "Japanese Monetary Policy, a Comparative Analysis." *Bank of Japan Monetary and Economic Studies* 4, pp. 105-127.

Eichenbaum, Martin. 1992. "Comments on ' Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy by Christopher Sims'." *European Economic Review* 36, pp. 1001-1011.

Feldman, R.A. 1986. *Japanese Financial Markets*. Cambridge, Mass: MIT Press.

Friedman, Benjamin. 1990. "Targets and Instruments of Monetary Policy." In *Handbook of Monetary Economics,* Volume II, Benjamin Friedman and Frank H. Hahn, eds. Amsterdam/New York: North Holland.

Friedman, Milton. 1985. "Monetarism in Rhetoric and in Practice." In *Monetary Policy in Our Times,* A. Ando, H. Eguchi, R. Farmer, and Y. Suzuki, eds., pp. 15-28. Cambridge, Mass: MIT Press.

Fukui, T. 1986. "Recent Developments of the Short-term Money Market in Japan and Changes in Monetary Control Techniques and Procedures by the Bank of Japan." Special Paper No. 130. The Bank of Japan Research and Statistics Department.

Hamada, K., and F. Hayashi. 1985. "Monetary Policy in Postwar Japan." In *Monetary Policy in Our Times,* A. Ando, H. Eguchi, R. Farmer, and Y. Suzuki, eds. Cambridge, Mass: MIT Press.

_____, and A. Horiuchi. 1987. "The Political Economy of the Financial Market." In *The Political Economy of Japan.* Vol. 1, K. Yamamura and Y. Yasuba, eds. Stanford: Stanford University Press.

Hutchison, Michael M. Forthcoming. "Structural Change and Macroeconomic Effects of Oil Shocks: Empirical Evidence from the United States and Japan." *Journal of International Money and Finance.* Issued as Working Paper No. PB92-06. Center for Pacific Basin Monetary and Economic Studies, Economic Research Department, Federal Reserve Bank of San Francisco.

_____. 1986. "Japan's 'Money Focused' Monetary Policy." Federal Reserve Bank of San Francisco *Economic Review* (Summer) pp. 33-46.

_____, and Carl Walsh. 1992. "Empirical Evidence on the Insulation Properties of Fixed and Flexible Exchange Rates: The Japanese Experience." *Journal of International Economics* 32, pp. 241-263.

Ito, T. 1982. "A Comparison of Japanese and U.S. Macroeconomic Behavior by a VAR Model." Discussion Paper No. 82-162. University of Minnesota. As reported in K. Okina, 1985.

_____. 1989. "Is the Bank of Japan a Closet Monetarist? Monetary Targeting in Japan, 1978-88." NBER Working Paper 2874.

Judd, J., and M. Hutchison. 1992. "Central Bank Secrecy and Money Surprises: International Evidence." *Review of Economics and Statistics* 74, pp. 135-145.

Kasman, B., and A. Rodrigues. 1991. "Financial Liberalization and Monetary Control in Japan." Federal Reserve Bank of New York *Quarterly Review* (Autumn).

Leeper, Eric M., and David B. Gordon. 1992. "In Search of the Liquidity Effect." *Journal of Monetary Economics* 29, pp. 341-369.

Meltzer, A. 1986. "Lessons from the Experiences of Japan and the United States under Fixed and Fluctuating Exchange Rates." *Bank of Japan Monetary and Economic Studies* 4(2) pp. 1-46.

Moreno, Ramon. 1992. "Are the Forces Shaping Business Cycles Alike? The Evidence from Japan." Working Paper No. PB92-10. Center for Pacific Basin Monetary and Economic Studies, Federal Reserve Bank of San Francisco.

Nakao, M., and A. Horii. 1991. "The Process of Decision-Making and Implementation of Monetary Policy in Japan." Special Paper No.198. The Bank of Japan Research and Statistics Department.

Ohkubo, T. 1983. "Money, Interest, Income and Prices." *Bank of Japan Monetary and Economic Studies* 1, pp. 111-146.

Okina, K. 1985. "Empirical Studies Using Granger Causality." *Bank of Japan Monetary and Economic Studies* 3, pp. 129-162.

Rudebusch, Glenn D. 1992. "Trends and Random Walks in Macroeconomic Time Series: A Re-examination." *International Economic Review* 33, pp. 661-680.

Sims, Christopher A. 1992. "Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy." *European Economic Review* 36, pp. 975-1000.

_____. 1986. "Are Forecasting Models Usable in Policy Analysis?" Federal Reserve Bank of Minneapolis *Economic Review* (Winter) pp. 2-16.

_____. 1980. "Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered." *American Economic Review* 70, pp. 250-257.

Strongin, Stephen. 1992. "The Identification of Monetary Policy Disturbances: Explaining the Liquidity Puzzle." Working Paper Series WP-92-27. Federal Reserve Bank of Chicago.

Suzuki, Y. 1986. *Money, Finance, and Macroeconomic Performance in Japan.* New Haven: Yale University Press.

_____. 1980. *Money and Banking in Contemporary Japan.* New Haven: Yale University Press.

_____, A. Kuroda, and H. Shirakawa. 1988. "Monetary Control Mechanism in Japan." *Bank of Japan Monetary and Economic Studies* 6, pp. 1-27.

Yasuda, T. 1981. "A Theoretical Interpretation of Window Guidance: A Game Theoretic Interpretation." Discussion Paper Series No. 4. Bank of Japan.

The Output-Inflation Trade-off in the United States: Has It Changed Since the Late 1970s?

John P. Judd and Jack H. Beebe

The authors are Vice President and Associate Director of Research, and Senior Vice President and Director of Research, respectively. They would like to thank Timothy Cogley, Frederick Furlong, Ramon Moreno, John Roberts, and Bharat Trehan for helpful suggestions on an earlier draft, and Sean Kelly and Andrew Biehl for research assistance.

In recent years, the Federal Reserve has become more explicit in stating a goal of gradually reducing inflation to near zero rates. An important consideration in seeking lower inflation is the transition cost (lost output and employment) incurred in the process. In this paper we ask whether the output-inflation trade-off in the U.S. is any more favorable now than it was in the high-inflation environment of the late 1970s and early 1980s. Our empirical estimates suggest that this trade-off is about the same as it was in the earlier period. In light of these results, we consider ways in which policies might be designed to reduce the amount of lost output associated with further disinflation.

Since late 1979, the Federal Reserve has pursued disinflationary monetary policies that can be characterized as occurring in two stages. First, in 1979–1981 the Fed successfully reduced inflation from double-digit to moderate rates of around 3½ percent in 1983–1985. Beginning in 1988, the Fed began explicitly stating that it intended to achieve a second period of disinflation, gradually moving the inflation rate from a moderate level of about 4 percent at that time to very low levels ("near" price stability) over a number of years. In 1992, CPI inflation was 3 percent, before dropping to about 2½ percent in the first ten months of 1993, indicating modest progress toward this goal.

An important consideration in seeking lower inflation is how to design policies that minimize the size of the transition costs that will be incurred in the process. These costs depend importantly on the credibility of the disinflation policy, i.e., on whether the public believes that the central bank actually will adhere to that policy. Thus a more (less) credible disinflation policy will translate more (less) quickly into lower inflation expectations, and therefore will have smaller (larger) effects on economic output.[1]

The costs associated with the policy of the early 1980s appeared to have been large, since the U.S. economy experienced the deepest recession of the post-World War II period in those years. This is not surprising. Over the prior decade the inflation rate had reached serious proportions, and thus the public may have needed to see some results before it began to believe in the Fed's resolve.

In this paper, we ask whether the transition costs have been any smaller in the recent disinflationary period than they were during the episode of the early 1980s. If so, it may be because the Fed's policies gained some credibility from its earlier disinflationary success, which reduced the

1. For an extensive review of the literature on monetary policy and policy credibility, see Blackburn and Christensen (1989). In their introduction, they note that ". . . the argument that figures prominently in contemporary discussions of deflationary management—namely that greater credibility of an anti-inflationary policy reduces the costs of disinflation—is persuasive" (p.2). Two approaches to designing an anti-inflation policy are discussed—gradualism, which implies a steady, predictable reduction in inflation, and immediacy, which aims at a more radical policy of cutting inflation more quickly. In this paper, we focus on the gradualist approach favored by the Fed.

size of the transition costs. If the costs were not smaller, then it may be because while the public believed that the Fed would not let inflation get out of control as in the late 1970s, the public was not convinced that it would reduce inflation from the moderate rates of the mid- to late-1980s to near zero.

We address this empirical issue by estimating the size of the short-run trade-off between output and inflation in the U.S. Our results suggest that this trade-off is about the same now as it was in the early 1980s. In addition, we point out that surveys of long-term inflation expectations suggest that the public expects inflation to rise a bit from present levels rather than decline according to the Fed's stated goal.

In light of these results, we consider ways in which policies could be designed to enhance credibility and thereby reduce the amount of lost output associated with a given amount of disinflation. First, and foremost, credibility is established through results: i.e., actually reducing the rate of inflation (Beebe 1991). However, it is possible that within the context of achieving a measure of success, lost output could be limited during disinflation if the Fed were more explicit about its disinflation goals. Thus having an explicit year-by-year inflation goal or range might help.

Going a step further, we also discuss the potential enhancements to credibility of finding an intermediate policy target to supplant the monetary aggregates, which have become less useful in recent years due to well-known instabilities. Consistently employing an intermediate target that is linked directly to the longer-term goal of reducing inflation might contribute to an expeditious enhancement of the credibility of that goal. Thus we suggest a class of intermediate-targeting approaches that might prove useful.

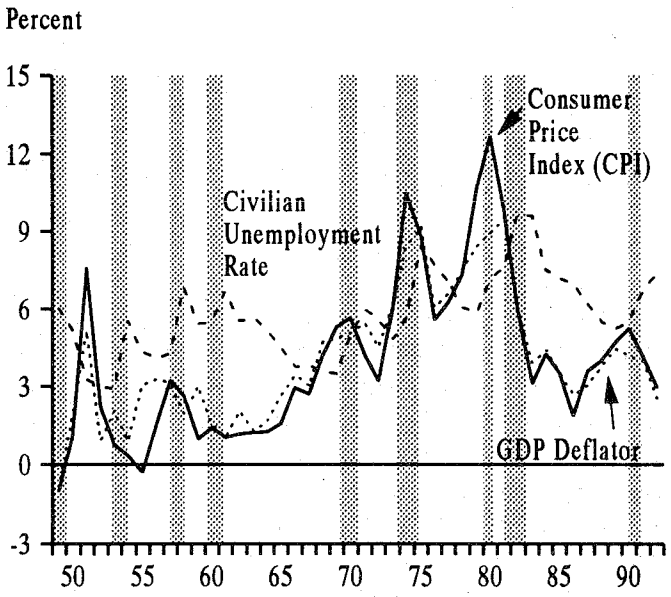
This paper follows with four sections. Section I is a brief discussion of Fed disinflationary policy since 1979. Section II provides evidence on the output-inflation trade-off. Section III provides evidence on long-term inflation expectations. Section IV offers suggestions on how credibility might be enhanced.

I. The Evolution of the Federal Reserve's Disinflationary Monetary Policy

By the time Paul Volcker became Federal Reserve Chairman in mid-1979, the expansionary monetary and fiscal policies of the late-1960s and 1970s had allowed consumer inflation to rise well into double digits (see Figure 1). These rates of inflation were very high by post-World War II standards and disrupted U.S. and world financial markets. U.S. long-term interest rates (for example, as measured by 20-year Treasury bond yields) rose from 4¼ percent in 1964 to 10¼ percent in late-1979, the dollar depreciated by nearly 25 percent between 1970–1979, and the price of gold rose to an historic high of over $800 an ounce in 1979 before settling back to over $400.

FIGURE 1

U.S. Inflation and Unemployment



Shaded areas represent recessions as defined by the NBER.
Inflation rates determined by a year/year calculation.

In response to these problems, the Fed dropped its practice of targeting the federal funds rate and instituted a new operating procedure under which it manipulated the quantity of reserves supplied to banks in an attempt to hit pre-announced ranges for several monetary aggregates. The main aggregate used was the narrow measure, M1, which includes currency in the hands of the public and fully checkable deposits. The new disinflationary policy consisted of attempting to achieve annual target ranges for M1, which would be gradually lowered over time.

The policy was successful in achieving its main goal: Between 1980 and 1983, CPI inflation fell from 12.7 to 3.1 percent (annual averages over the prior year). The cost was the most severe recession in post-World War II history, in which the civilian unemployment rate peaked at 10.8 percent in late-1982 and averaged over 9½ percent in both 1982 and 1983 (Figure 1).

By 1983 the operating procedures of monetary policy had shifted. First, the Fed de-emphasized M1 in favor

of a broader aggregate, M2. Problems with M1 appear to have stemmed from both financial innovation and deregulation. Such new instruments as repurchase agreements and money market mutual funds were close substitutes for the deposits in M1, and therefore led to instability in its velocity. The availability of these new instruments was a major impetus behind the removal of deposit interest rate ceilings, mainly from 1978 to 1983. However, deregulation also created problems by blurring the distinction between transactions and savings balances held at depository institutions. The rationale for emphasizing M2 was that it was broad enough to internalize much of the portfolio substitution that had disrupted M1.

Second, in day-to-day operations, the Fed began to focus on the quantity of reserves borrowed from Reserve Banks as its operating instrument, which is similar to using the federal funds rate as the instrument of policy (Wallich 1984). Moreover, the degree of precision in monetary targeting was reduced, and money once again became one among a number of important indicators for policy, including data on developments in the real economy and prices, as well as in the domestic and international financial markets (Heller 1988).

The explicitness of a "price stability" goal did not appear until late in the 1983-90 expansion. Early in the expansion, official statements of Chairman Volcker generally were vague as to an inflation goal. For example, the February 6, 1984 *Monetary Report to Congress* stated that, "The (monetary) ranges for 1984 are intended to be consistent with the basic objective of achieving long lasting economic expansion in a context of continuing control of inflationary pressures."

However, after becoming Chairman in 1987, Mr. Greenspan stated explicitly in his monetary reports to Congress that the Fed's long-term goal was price stability, although neither a time frame nor specific annual goals for inflation were established. In his testimony of February 23, 1988, accompanying his first monetary report to Congress, Chairman Greenspan stated, "Progress toward price stability is the foundation on which the longest peacetime expansion in our nation's history has been built, and continued efforts along this line will be the framework for future economic advances." The February 20, 1990 *Monetary Report to Congress* stated that, "The Federal Open Market Committee is committed to the achievement, over time, of price stability." Moreover, Chairman Greenspan and a number of Reserve Bank Presidents supported a bill introduced by Congressman Neal requiring the Fed to achieve price stability within five years. Given the focus of the Greenspan Fed on price stability, it may be instructive to think of two disinflationary sub-periods since 1979: the early 1980s in which inflation was reduced to around 3½ percent, and the period since the late-1980s in which a further reduction has been sought.

Over the 1983–1990 expansion, little or no progress was made in reducing inflation below the 4 percent rate that had been established by 1984. In fact, following temporarily low inflation in 1986, caused by a sharp drop in the oil price, inflation began rising somewhat again, reaching over 5 percent on a consumer price basis by 1990 (although the latter rate was boosted by a temporary surge in the oil price.)

By mid-1989, the U.S. economy had slowed substantially, growing at less than a 2 percent rate, until it fell into recession in mid-1990. The recession, which was relatively mild and lasted three quarters, was followed by a long period of slow, but positive, growth in 1991 through mid-1993. In response to the overall pattern of slow economic growth since 1989, inflation has shown signs of a downward trend, averaging about 3 percent in 1992 and 2½ percent over the first ten months of 1993.

Since the onset of the 1990–1991 recession, Fed policy has focused on boosting economic growth moderately, although it has retained its long-run goal of gradually reducing inflation to very low levels. M2 growth has remained extremely weak. Especially in 1991 and 1992, M2 came in near the bottom of, or below, its annual range. However, low M2 growth has not been considered a reliable measure of monetary tightness, since M2's relationship to other economic variables appears to have shifted significantly. Like the earlier problem with M1, the problem with M2 seems to have arisen primarily from financial deregulation and innovation (Judd and Trehan 1992).

For these reasons, M2 has been de-emphasized in policy decisions by the Fed in recent years. In essence, the Fed has not had any monetary aggregate considered reliable enough to use as a primary guide to monetary policy. Instead, in recent years, it has relied on purely discretionary adjustments to the federal funds rate to find a delicately balanced policy geared toward promoting moderate economic growth, while making further progress in reducing inflation.

II. Empirical Evidence on the Output-Inflation Trade-off

In this section, we assess whether the output-inflation trade-off has shifted downward since the late 1970s, when the Fed increased its emphasis in public statements and actions on the goal of reducing inflation. To do so, we analyze an equation commonly used to estimate the trade-off, and we review movements in inflation expectations as measured by surveys.

To estimate the trade-off, we use:

$$(1)\ \Delta p_t = \alpha + \lambda \Delta x_t + \beta \Delta p_{t-1} + \gamma(y_{t-1} - \bar{y}_{t-1})$$

where, $x_t \equiv p_t + y_t$; x = log of nominal GDP; p = log of aggregate price level; y = log of real GDP; and $\bar{y}$ = log of trend real GDP.

This equation has been used to estimate the trade-off by authors with such diverse views about the structure of the economy as Lucas (1973), Gordon (1983), Gordon and King (1982), Schultze (1984), and Ball, Mankiw, and Romer (1988).[2] Thus equation (1) appears to be consistent with both "new" and "old" Keynesian theory as well as demand-oriented, or monetarist, equilibrium business cycle theory. Correlations of the type expressed by the equation should be evident in both (1) an economy in which expectations are adaptive, so that an expectations-augmented Phillips-curve would apply, and (2) an economy in which expectations are rational, so that the "trade-off" represents only an observed short-run correlation that is not exploitable by policymakers. Thus Lucas (1973) derives a relationship like equation (1) from a monetary-misperceptions model with rational, optimizing agents, and Gordon (1983) shows how equation (1) can be viewed as a rearranged version of an adaptive-expectations Phillips curve.[3]

The key assumption underlying the equation is that growth in nominal GDP is exogenous with respect to inflation. As such, it would capture the effects of aggregate demand on inflation, and would be independent of aggregate supply shocks. (The viability of this assumption is assessed below.) Then, for a given lagged inflation rate and state of the business cycle, the coefficient λ measures the proportion of the change in aggregate demand that affects prices in the short-run as opposed to output. The output-inflation trade-off is calculated as $\tau = (1-\lambda)/\lambda$. It measures the percentage point change in output per percentage point of change in inflation resulting from a given change in aggregate demand. If the Fed's disinflation policy has gained credibility over the 1980s, then λ should have risen and τ should have declined over this period. Other coefficients in the equation also might have changed. However, following the earlier literature, we will focus exclusively on λ and its implications for τ.

Estimating the Trade-off

Table 1 presents the results of estimating various forms of equation (1) using annual data over 1949 to 1992. The simplest estimated equation is shown in column 1. In this column, the cyclical variable $(y-\bar{y})$ is formed by linearly de-trending real GDP. (An alternative de-trending method is discussed below.) All of the explanatory variables have the expected signs, are highly significant, and together account for about 82 percent of the variation in annual inflation. This regression suggests that τ averaged about 1.7 (.63/.37) in the U.S. in the post-World War II period.[4]

However, in order to feel comfortable with the assumption that x is exogenous with respect to p, it is necessary to investigate the issue of whether supply shocks are likely to be biasing estimates of λ. A supply shock causes p and y to move in opposite directions. If these variables do *not* move by equal (proportional) amounts, then there will be a resulting movement in nominal GDP, which will produce a correlation between nominal GDP and inflation that would be misinterpreted by the equation as reflecting the trade-off. In other words, supply shocks will bias estimates of λ unless the aggregate demand function has a (negative) unitary elasticity.

Columns 2 through 6 represent attempts to see if supply shocks present a problem in estimating λ. First, we introduce supply shock variables to see if the estimate of λ is altered substantially. Second, we use two-stage least squares estimation to eliminate possible reverse causation, and again observe whether this affects the estimate of λ. In column 2, we add a dummy variable that attempts to capture the effects of major oil shocks, by taking on the value of 1 in 1974 and 1979 and -1 in 1986. This variable is significant in the equation and has the expected sign, but does not significantly alter the estimate of λ. Column 3 shows two-stage-least squares estimates of the same equation that was estimated with OLS in column 2. Again, the estimate of λ is not materially affected. Column 4 introduces changes in the relative price of oil (from the Producer Price Index), and has no effect on λ. In column 5, we test

2. Other papers dealing with this issue are Ball (1991, 1993), Friedman (1984, 1988), and Okun (1978).

3. In Judd and Beebe (1993, pp. 306 and 317), we tested for the stability of an inflation-augmented Phillips curve, which expressed wage inflation as a function of slack in the labor market (as measured by the unemployment rate relative to its estimated full employment level), and expected wage inflation (as measured by past wage inflation). These tests can be considered an alternative way of testing for the stability of the inflation-unemployment trade-off. Similar to the results discussed below for equation 1, we failed to reject stability of the wage inflation Phillips curve.

4. Using monte carlo methods, we calculated the t statistic for the test of whether this estimate of τ is different from zero, based upon the estimate of λ and its standard error in column 1. The t statistic for τ was estimated to be 4.33, suggesting that τ is different from zero with a high level of confidence.

TABLE 1

OUTPUT-INFLATION TRADE-OFF EQUATIONS: ALTERNATIVE SPECIFICATIONS
1949–1992

	(1)	(2)	(3)[1]	(4)	(5)[2]	(6)[3]
Constant	−1.07**	−0.96**	−0.95**	−0.91**	−0.99**	−0.015**
Δx_t	0.37**	0.36**	0.37**	0.36**	0.37**	0.42**
Δp_{t-1}	0.44**	0.41**	0.42**	0.40**	0.41**	0.66**
y_{t-1}	0.15**	0.13**	0.13**	0.12**	0.14**	—
T_t	−0.0041**	−0.0036**	−0.0036**	−0.0034**	−0.0037**	—
$y_{t-1} - \bar{y}_{t-1}$	—	—	—	—	—	0.41**
$D0$	—	0.013*	0.012*	—	0.011*	—
$\Delta poil_t$	—	—	—	0.030*	—	—
$\Delta forex_t$	—	—	—	—	0.00013	—
$\bar{R}^2$	0.82	0.84	0.84	0.85	0.83	0.83
SEE	0.010	0.0099	0.0099	0.0096	0.010	0.010
Q(11)	8.08	7.39	7.59	8.52	8.32	6.61

NOTE: Marginal significance levels: * = .05; ** = .01.

1. This equation was estimated with a two-stage least squares procedure. Instrumental variables used for Δx_t include Δx_{t-1}, Δb_{t-1}, ΔDEF_t, Δp_{t-1}, y_{t-1}, T_t, $D0$.

2. This equation was estimated with a two-stage least squares procedure. Instrumental variables used for Δx_t and $\Delta forex_t$ include Δx_{t-1}, Δb_{t-1}, ΔDEF_t, Δp_{t-1}, y_{t-1}, T_t, $D0$ and $\Delta forex_{t-1}$.

3. $\bar{y}$ was estimated as the permanent component of y from a VAR for y and the six-month commercial paper rate as in Judd-Trehan (1990). Because of well-known problems associated with using generated regressors (Pagan 1984), the t statistic on $\bar{y}_{t-1}$ in column 6 is biased upward.

Definition of variables:

x = log of nominal GDP $\equiv p + y$
p = log of GDP deflator
y = log of real GDP
$\bar{y}$ = log of trend real GDP
T = time
$poil$ = log of relative price of energy, producer price index
$forex$ = real trade-weighted exchange rate beginning in 1969, zero from 1949 to 1968

$$D0 = \begin{cases} 1 \text{ in } 1974, 1979 \\ -1 \text{ in } 1986 \\ 0 \text{ elsewhere} \end{cases}$$

b = log of monetary base (FRB St. Louis)
DEF = log of nominal federal defense expenditures

for possible effects of changes in the real trade-weighted value of the dollar. Using two-stage-least squares methods, we again find no significant effect on the estimated size of λ. We conclude from these exercises that the basic equation does not appear to be distorted by the effects of supply shocks.

A second issue in estimating λ has to do with how to de-trend real GDP to form the business cycle variable (see Rudebusch 1993). In the estimated equations discussed above, we used a linear time trend to represent equilibrium real GDP. In a second somewhat more complex approach, we used the method of Blanchard-Quah (1989) to extract the trend component. This method involves estimating a structural VAR with the identifying restriction that there are two types of shocks—a permanent and a transitory shock.[5] The permanent shock is associated with trend real output, while the transitory shock is associated with the business cycle. Thus, we introduced the transitory component of real GDP, as estimated by this method, into the equation in column 6 in place of y_{t-1} and T in column 1. Based upon the estimates of λ in columns 1 and 6, this substitution reduced the estimate of τ by 19 percent ((1.70 − 1.38)/1.70). In the discussion below, we test for possible shifts in λ using both methods of de-trending y, to be sure that this factor does not affect our results.

Tests for Shifts in the Trade-off

In Table 2, we present tests for shifts in λ. First, we take columns 2 and 6 in Table 1 and introduce a dummy variable times the growth in nominal GDP, which yields columns 7 and 9. These latter columns provide a test for a decline in λ over 1980–1992 compared with 1949–1979. Column 7 (like column 2) uses linearly de-trended real GDP while column 9 (like column 6) uses the Blanchard-Quah method of de-trending. In both equations, the estimated λ rises somewhat (from .36 in 2 to .42 in 7 and from .42 in 6 to .46 in 9), but neither change is statistically significant even at the 10 percent level.

Using monte carlo methods, we calculated the t statistics for a change in τ in 1980–1992 based upon the estimates of λ and their standard errors in columns 7 and 9. The results were the same qualitatively as those for λ: We were not able to reject stability even at the 10 percent level.

One potential problem with the tests in columns 7 and 9 is that the period from 1949 to 1979 encompasses years in which inflation was low (1949–1965), as well as years in which inflation increased (1965–1979). The tests in column 7 and 9 ask whether λ was different in 1980–1992 from the *average* ratio in the *entire* prior period, whereas we are more interested in seeing if it rose in 1980–1992 compared with the period in which inflation rose (1965–1979). Columns 8 and 10 attempt to address this question by including slope dummy variables (on Δx) for 1965–1992 and for 1980–1992. Although column 8 shows a decrease in λ beginning in 1965, neither column 8 nor 10 suggests a significant shift since the late 1970s.[6]

As a final check, we consider the possibility that λ may have changed gradually following the late 1970s as the public learned of the Fed's increased resolve to reduce inflation. In Table 3 (p. 32) we test for a shift in λ in blocks stretching from *each year* in 1980–1992 to the end of the sample. Again, we do not find any single dividing point in which there is a significant change in λ, even at the 10 percent level of significance.

In summary, despite a considerable search for a shift in λ after the late 1970s, we have found none. It appears that the Fed faces about the same output-inflation trade-off today in attempting to reduce inflation from its present moderate level that it faced at the height of the inflation and financial instability in 1979.

III. INFLATION EXPECTATIONS

Our conclusion that the output-inflation trade-off has not shifted seems consistent with the evidence from surveys of inflation expectations, which have been slow to adjust to disinflationary results. As shown in Figure 2 (p. 32), expectations in 1980 through 1982 of average inflation over the next ten years were well above subsequent actual ten-year average inflation rates for the ten-year-ahead period. Even by 1982, average inflation expected over the next 10 years was 6¾ percent, while the *ex post* realized average turned out to be only 4 percent.

A decade later, long-run inflation expectations remain well above the 1992 inflation rate of around 3 percent. Financial decisionmakers, as represented by the

5. Following Judd-Trehan (1990), we estimated a two-variable VAR for log changes in real GDP and the change in the commercial paper rate, using six lags of each variable. This system yielded impulse response functions similar to those commonly found in the literature. Thus, for example, positive transitory (demand) shocks cause output to rise temporarily before returning to trend, while positive permanent (supply) shocks cause output to rise permanently.

6. Ball, Mankiw, and Romer (1988) argue that the sacrifice ratio should rise (fall) as inflation falls (rises) because of menu costs, and they present cross-sectional evidence from a number of different countries that such a relationship exits. However, using time-series data, Ball (1993) fails to find this effect. We tested for this effect by including both Δx and Δx^2 in regressions 7 and 9, Table 2. The combination of these two variables means that the sacrifice ratio can vary with the growth rate of nominal demand. Like Ball, we failed to find a significant effect in our time-series data.

TABLE 2

TESTING FOR CHANGES IN THE OUTPUT-INFLATION TRADE-OFF BEGINNING IN 1965 AND 1980

	(7)	(8)	(9)	(10)
Constant	−1.08**	−1.56**	−0.015**	−0.014**
Δx_t	0.35**	0.48**	0.43**	0.41**
$\Delta xD65_t$	—	−0.22**	—	0.02
$\Delta xD80_t$	0.073	0.075	0.026	0.022
Δp_{t-1}	0.41**	0.49**	0.65**	0.65**
y_{t-1}	0.15**	0.21**	—	—
T_t	−0.0043*	−0.0059**	—	—
$y_{t-1} - \bar{y}_{t-1}$	—	—	0.43**	0.42*
$D0$	0.014*	0.013*	—	—
$\bar{R}^2$	0.84	0.86	0.81	0.81
SEE	0.001	0.009	0.010	0.010
Q(11)	9.09	6.79	6.03	6.90

NOTE: Marginal significance levels: * = .05; ** = .01.

Definition of variables:

$$D65 = \begin{cases} 1 & 1965\text{–}1992 \\ 0 & \text{elsewhere} \end{cases}$$

$$D80 = \begin{cases} 1 & 1980\text{–}1992 \\ 0 & \text{elsewhere} \end{cases}$$

Other variables are defined in Table 1.

Hoey/Philadelphia Fed survey, expect inflation to average 3¾ percent over the next ten years (survey of 1993.Q2), while the Michigan survey suggests that households expect a 5¼ percent average inflation rate over the same period (average of January through May 1993 surveys). It appears that the public remains unconvinced that the Fed will achieve inflation much below 4 percent, despite the stated goal of price stability.

IV. WHAT CAN BE DONE TO MAKE DISINFLATION LESS COSTLY?

Apparently, the Fed faces the same output-inflation trade-off now that it faced in the early 1980s as it sought to bring inflation down from double-digit rates. Of course, one reason for this may be that it has yet to produce clear results in reducing inflation significantly below the level established in the mid-1980s. Although it seems unlikely that the public fears another outbreak of double-digit inflation, evidence has not yet been observed supporting the view that inflation will move to a level much below those that have prevailed since the mid-1980s.

Doubts about lower inflation may be magnified by large actual and projected federal budget deficits since the early 1980s. There may be concern that in the long run, persistently large deficits will lead to higher inflation, even though the Fed generally is credited with not having succumbed to pressure to monetize the federal debt to date.

In addition, the current design of monetary policy may not make it easy for the public to discern how much emphasis is being placed on inflation reduction. Although the Fed has stated for a number of years that its main objective is to eliminate inflation, it also has paid attention to output stabilization. The expressed intent of mitigating

TABLE 3

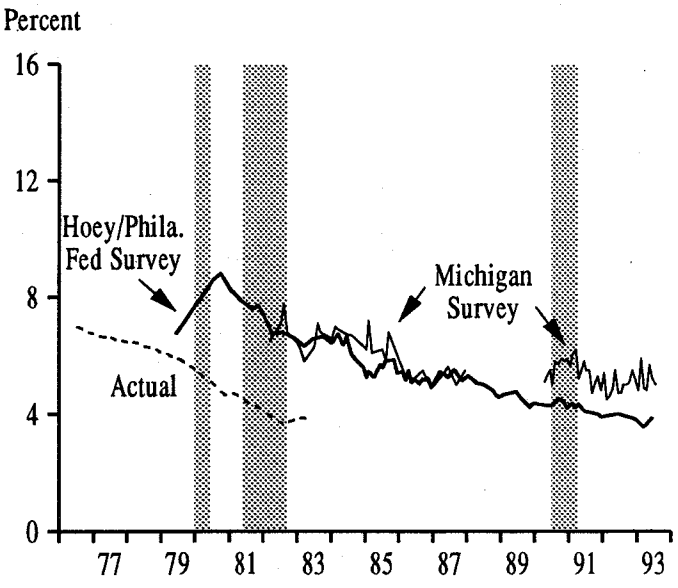
TESTING FOR CHANGES IN THE OUTPUT-INFLATION TRADE-OFF IN 1980 THROUGH 1992

	CHANGE IN COEFFICIENT ON Δx	
PERIOD TESTED	(8)	(10)
1980–92	.076 (0.25)	.022 (0.68)
1981–92	.021 (0.76)	−.006 (0.91)
1982–92	−.121 (0.18)	−.049 (0.47)
1983–92	−.117 (0.20)	−.048 (0.47)
1984–92	−.071 (0.43)	−.041 (0.55)
1985–92	−.101 (0.34)	−.043 (0.59)
1986–92	−.038 (0.73)	−.017 (0.84)
1987–92	−.048 (0.64)	−.002 (0.98)
1988–92	−.022 (0.84)	−.025 (0.80)
1989–92	.006 (0.96)	−.024 (0.84)
1990–92	.019 (0.91)	.009 (0.96)
1991–92	.052 (0.81)	.011 (0.96)
1992	−.067 (0.76)	−.014 (0.55)

NOTE: Marginal significance levels shown in parentheses. For example, a marginal significance level of 0.10 would suggest that stability could be rejected with the probability of 10 percent that the equation is stable. The power of these tests declines as the dividing point in the sample moves toward the end of the sample.

FIGURE 2

EXPECTED AND ACTUAL CPI INFLATION FOR TEN YEARS AHEAD



Shaded areas represent recessions as defined by the NBER. Expected and actual data are for averages over the next 10 years, not for the 10th year out.

cyclical downturns inevitably raises the issue of whether this goal will take precedence over disinflation at any particular time. Given the discretionary approach followed by the Fed, in which it resolves conflicts between the two goals on a case-by-case basis, it may be difficult for the public to be sure that the Fed's resolve to reduce inflation has not flagged.

One approach that might help convince the public that the Fed is serious about disinflation would be to announce specific inflation targets, or at least target ranges, for the years ahead.[7] By showing a willingness to commit itself to a particular path of disinflation, and thereafter, to a particular range for inflation, the Fed might be making its resolve more credible. Moreover, it would be providing the market with a benchmark for judging progress in meeting that goal.

A related issue concerns intermediate targets for monetary policy. As discussed above, although the Fed establishes target ranges for the monetary aggregates, it often does not take actions to achieve those targets, since rapid financial change has made it inadvisable to adhere to rigid targets for these variables. As a consequence, however, the market has received ambiguous and confusing signals about what the Fed is doing to achieve its long-run disinflation goal. If the Fed had been able to pursue its monetary target variables more aggressively, it might have enhanced the credibility of its disinflation goal by providing the market with timely feedback on whether it was acting in the short run in a way that would achieve its long-run inflation goal (Cukierman-Meltzer 1986).

Given the problems with the monetary aggregates, it seems worthwhile to evaluate the usefulness of alternative intermediate target variables and targeting procedures. Recent research outlined briefly in the accompanying box suggests that nominal GDP possibly could be used effectively as an intermediate target in a context in which the Fed retains its use of a nominal interest rate as its instrument of policy (Judd and Motley 1992 and elsewhere in this *Review*). Essentially, the approach involves raising (lowering) a short-term nominal interest rate whenever growth in last period's nominal GDP exceeds (falls short of) a preestablished target for nominal GDP growth. The targeted growth rate for nominal GDP would be chosen to be consistent with a goal for inflation and made explicit *ex ante*.

A monetary policy rule such as this offers several potential advantages. First, nominal GDP would not be disrupted by shifts in the velocity of money. Second, to construct such a rule, the Federal Reserve would first have to specify an inflation goal. Moreover, by linking specific

7. In this discussion, we confine ourselves to ways of improving the credibility of gradualist disinflation policies. For a discussion of the merits of "cold-turkey" approaches, see Ball (1993) and Sargent (1983).

LINKING AN INFLATION GOAL TO AN INTERMEDIATE TARGET AND OPERATING PROCEDURE

The following rule is used to illustrate an approach to policy that might have advantages from the point of view of expeditiously establishing credibility. The feedback rule links movements in a short-term interest rate to nominal GDP:

$$\Delta R_t = \delta[\Delta x_{t-1} - \Delta x^*_{t-1}].$$

The variable R denotes the policy instrument, which in this case is a short-term nominal interest rate such as the federal funds rate that is under the direct *short-run* control of the monetary authority. The variable x represents the intermediate target variable of policy, which in this case is (the log of) nominal GDP. The rule specifies that the change in the interest rate each quarter is a function of last quarter's deviation between the growth rate of nominal GDP (Δx) and its target growth rate (Δx^*).

The targeted growth rate of nominal GDP would be set according to:

$$\Delta x^*_t = \Delta p^*_t + \Delta \bar{y}_t,$$

where Δp^* is the central bank's inflation target and $\Delta \bar{y}$ is the estimated trend growth rate of real GDP. The strength of the monetary authority's response to deviations between Δx and Δx^* is defined by δ, and can be selected by the central bank. Based upon stochastic simulations of two small macroeconomic models, this rule appears to be capable of holding long-run inflation to within fairly narrow bounds, without substantially increasing volatilities in real GDP or interest rates above those observed in the post-war period (Judd-Motley 1992 and this issue.)

policy actions (i.e., changes in a short-term interest rate) to an intermediate target that is simply and clearly linked to the inflation goal, the public would have a simple way to monitor the Fed's resolve to achieve and maintain that inflation goal. Finally, the rule *either* could be followed explicitly by the Fed or be used to guide and assess a discretionary policy, should the Fed wish to diverge from the policy prescribed by the rule. While a full assessment of such an approach would involve issues other than credibility, it appears that an approach of this type, whether used as a rule or as a baseline for discretion, might reduce the cost of disinflation.

In conclusion, the empirical tests in this paper suggest strongly that the output cost of reducing inflation is about the same as it was at the height of the inflationary period from the late 1970s to the early 1980s. It is possible, however, that this cost might be reduced if the Fed were to make a public commitment to an explicit inflation target and perhaps if it also were to commit itself to an intermediate target and operating procedure linked explicitly to the inflation target.

REFERENCES

Ball, Laurence M. 1993. "What Determines the Sacrifice Ratio?" Paper presented at the Conference on Monetary Policy, January 21-24, Islamorada, Florida, National Bureau of Economic Research.

_____. 1991. "The Genesis of Inflation and the Costs of Disinflation." *Journal of Money, Credit and Banking* (August, Part 2) pp. 439-452.

Ball, Laurence N., Gregory Mankiw, and David Romer. 1988. "The New Keynesian Economics and the Output-Inflation Trade-off." *Brookings Papers on Economic Activity,* pp. 1-82.

Beebe, Jack. 1991. "Comments on 'Japanese Monetary Policy during 1970-90: Rules or Discretion?' by Kazuo Ueda." *Proceedings of the Bank of Japan Monetary Conference,* October 24-25.

Blackburn, Keith, and Michael Christensen. 1989. "Monetary Policy and Policy Credibility: Theory and Evidence." *Journal of Economic Literature* (March) pp. 1-45.

Blanchard, Olivier J., and Danny Quah. 1989. "The Dynamic Effects of Aggregate Demand and Supply Disturbances." *American Economic Review* (September) pp. 655-673.

Board of Governors of the Federal Reserve System. 1984, 1988, 1990. *Monetary Report to Congress Pursuant to the Full Employment and Balanced Growth Act of 1978,* February 6, February 23, and February 20, respectively.

Cukierman, Alex, and Allan H. Meltzer. 1986. "The Credibility of Monetary Announcements." In *Monetary Policy and Uncertainty,* ed. Manfred Neuman. Baden-Baden: Nomos Verlagsgesellschaft GW.

Friedman, Benjamin. 1988. "Lessons on Monetary Policy from the 1980s." *Journal of Economic Perspectives* (Summer) pp. 51-72.

_____. 1984. "Lessons from the 1979-82 Monetary Policy Experiment." *American Economic Review (Papers and Proceedings)* (May) pp. 382-387.

Gordon, Robert J. 1983. "A Century of Evidence on Wage and Price Stickiness in the United States, the United Kingdom, and Japan." In *Macroeconomics, Prices and Quantities,* pp. 85-133. Washington: The Brookings Institution.

_____, and Stephen R. King. 1982. "The Output Cost of Disinflation in Traditional and Vector Autoregression Models." *Brookings Papers on Economic Activity* pp. 205-244.

Heller, Robert H. 1988. "Implementing Monetary Policy." *Federal Reserve Bulletin* (July) pp. 419-429.

Judd, John P., and Jack H. Beebe. 1993. "The Credibility of the Federal Reserve's Disinflation Policy in a Period of Financial Change." In *Financial Deepening and Economic Growth in the Pacific Basin Economies,* eds. Yang Wenyou and Zhang Fengming, pp. 287-317. Foreign Affairs Department, People's Bank of China.

_____, and Brian Motley. 1992. "Controlling Inflation with an Interest Rate Instrument." Federal Reserve Bank of San Francisco *Economic Review* 3, pp. 3-22.

_____, and _____. 1993. "Using a Nominal-GDP Rule to Guide Discretionary Monetary Policy." Federal Reserve Bank of San Francisco *Economic Review,* this issue.

_____, and Bharat Trehan. 1992. "Money, Credit and M2." Federal Reserve Bank of San Francisco *Weekly Letter* (September 4).

_____, and _____. 1990. "What Does Unemployment Tell Us about Future Inflation?" Federal Reserve Bank of San Francisco *Economic Review* (Summer) pp. 15-25.

Lucas, Robert E., Jr. 1973. "Some International Evidence on Output-Inflation Tradeoffs." *American Economic Review* (June) pp. 326-334.

Okun, Arthur M. 1978. "Efficient Disinflationary Rules." *American Economic Review* (May) pp. 348-352.

Pagan, Adrian R. 1984. "Econometric Issues in the Analysis of Regressions with Generated Regressors." *International Economic Review* (February) pp. 221-247.

Rudebusch, Glenn D. 1993. "The Uncertain Unit Root in Real GNP." *American Economic Review,* forthcoming. (Currently available in Finance and Economics Discussion Series, Number 193, Board of Governors of the Federal Reserve System, April 1992).

Sargent, Thomas J. 1983. "Stopping Moderate Inflations: The Methods of Poincare and Thatcher." In *Inflation, Debt, and Indexation,* eds. R. Dornbusch and M. Simonsen pp. 54-95. Cambridge, Mass: MIT Press.

Schultze, Charles L. 1984. "Cross-Country and Cross-Temporal Differences in Inflation Responsiveness." *American Economic Review (Papers and Proceedings)* (May) pp. 160-165.

Wallich, Henry. 1984. "Recent Techniques of Monetary Policy." Federal Reserve Bank of Kansas City *Economic Review* (May) pp. 21-30.

Adapting to Instability in Money Demand: Forecasting Money Growth with a Time-Varying Parameter Model

Timothy Cogley

Senior Economist, Federal Reserve Bank of San Francisco.
I am grateful to Hopi Huh, Mark Levonian, and Bharat Trehan for helpful comments on an earlier draft and to Desiree Schaan for computational assistance.

Conventional money demand models appear to be unstable, and this complicates the problem of conducting monetary policy. One way to deal with parameter instability is to learn how to adapt quickly when parameters shift. This paper applies a time-varying-parameter estimator to conventional money demand models and evaluates its usefulness as a forecasting tool. In relative terms, the time-varying-parameter estimator improves significantly on ordinary least squares. In absolute terms, we continue to have difficulty tracking money demand through turbulent periods.

According to the quantity theory of money, nominal spending depends on the supply of money and on velocity, and velocity is determined by money demand. If money demand is stable, monetary aggregates can be used as indicators of fluctuations in nominal aggregate demand. Furthermore, if money demand is functionally invariant to changes in money supply, then the Federal Reserve may be able to adjust the money supply in order to offset fluctuations in nominal spending that are due to non-monetary disturbances.

Conventional models of money demand appear to be unstable, however, and this greatly complicates the problem of conducting monetary policy. In particular, since money demand models are functionally unstable, it is difficult to interpret the information in monetary aggregates. For example, in recent years, the Federal Reserve System's money demand models have consistently underestimated M2 velocity. As a consequence, the Federal Reserve has overestimated the rate of M2 growth needed to sustain the projected growth in nominal GDP, and therefore actual M2 growth has fallen below its target range. Ordinarily, the unexpected shortfall in M2 growth would be a sign of serious weakness in the economy. However, in this case, it simply reflected the fact that velocity turned out to be higher than expected. Thus instability in money demand models makes it difficult for the Federal Reserve to keep money within its target range while still trying to achieve its goals for the economy.

As a theoretical matter, there is no reason to believe that conventional money demand models should be stable. For example, since conventional representations are subject to the Lucas critique, changes in central bank operating procedures can alter money demand parameters. Similarly, financial innovation may alter the relation between velocity and opportunity costs. Thus it seems appropriate to treat conventional money demand models as time-varying parameter models.

Roughly speaking, there are two ways to deal with time-varying parameters. One is to seek a deeper theoretical structure whose parameters are time invariant. So far, monetary economists have had little success with this approach. Another way to deal with parameter instability is to learn how to *adapt* to functional changes

in money demand by allowing estimated parameters to change quickly when the model begins to show signs of instability. This paper takes the latter approach. It explores a time-varying parameter estimator that gives more weight to recent data and less weight to older data, so that estimates can change quickly when parameters change. The goal is to improve the predictive performance of money demand models.

This intuition is formalized in terms of discounted least squares (DLS). The paper applies recursive DLS to a number of conventional money demand models and compares its predictive performance with ordinary least squares (OLS). In relative terms, DLS compares favorably to OLS. For example, in cases where instability is especially important, DLS reduces the mean square error of one-quarter-ahead forecasts by 55 to 60 percent. Thus DLS can provide an important hedge against gross instability.

In absolute terms, however, conventional money demand models still have a great deal of trouble forecasting through turbulent periods. Thus DLS represents only a partial solution to parameter instability. In particular, since we continue to have difficulty tracking M2 demand, it will continue to be difficult to use M2 as an indicator of economic conditions.

I. TIME-VARYING PARAMETERS

This section interprets conventional money demand functions in order to motivate the empirical approach taken in the paper. In conventional money demand models, demand for real balances depends on a scale variable, such as income, consumption, or wealth, and on opportunity cost variables. For example, Meltzer (1963) studied variations on the following model:

$$ln\,(m_t / p_t) = \beta_0 + \beta_1\, ln\,(r_t) + \beta_2\, ln\,(w_t) + u_t,$$

where m_t denotes nominal money balances, p_t is the price level, r_t is a nominal interest rate, w_t is either real wealth or income, and β_0, β_1, and β_2 are parameters.

Most of the empirical literature assumes that the parameters are *time invariant*. In practice, however, estimated money demand models appear to be unstable. For example, there was the famous "case of the missing money" in the mid-1970s (see Goldfeld 1976, or Judd and Scadding 1982). More recently, Feinman and Porter (1992) report evidence that M2 demand models have gone off course.

From a theoretical point of view, the instability of empirical money demand models is not puzzling. On the contrary, it is exactly what monetary theory predicts. Two kinds of arguments generate time-varying parameters, one based on the Lucas critique and another based on financial innovation.

The first argument concerns identification and is due to Cooley and LeRoy (1981). Traditional money demand models describe an equilibrium between money supply and money demand. In order to interpret the parameters solely in terms of money demand, however, money supply must be predetermined or exogenous. This condition seems dubious. If money supply is endogenous, the estimated parameters will depend at least in part on supply factors. Furthermore, if there are changes in the determinants of money supply, such as a change in monetary policy operating procedures, the parameters of conventional money demand models will also change. Thus, when money supply is endogenous, a necessary condition for parameter stability is that monetary policy rules not change during the sample. Since post-war U.S. data probably do not satisfy this condition, parameter instability is to be expected.

A second argument emphasizes financial innovation (e.g., Ireland 1992). For example, in cash-in-advance models, agents can buy some goods only with money and other goods with either money or credit. The cash-in-advance constraint gives rise to a transactions demand for money. A financial innovation expands the set of goods that can be bought on credit and thus (other things equal) reduces demand for real balances. Thus financial innovation alters the relation between money, interest rates, and expenditures. Since conventional money demand models do not fully capture the effects of financial innovation, one should expect parameter instability during periods of financial innovation.

Recently, a number of authors have argued that financial innovation may account for the recent bout of instability in M2 demand. In particular, the increased availability of mutual funds may have altered the relation between M2 velocity and interest rates (e.g., Feinman and Porter 1992 or Duca 1993). For example, banks have begun to market mutual funds to retail customers. As these funds become more accessible, the transactions costs of switching between M2 and various non-M2 securities are reduced. This increases the substitutability between M2 and stock and bond funds and thus increases the interest elasticity of M2 demand. The unusually steep yield curve of the last few years also may have induced some investors to switch into mutual funds.

These arguments suggest that we should treat conventional money demand functions as time-varying parameter models. Broadly speaking, there are two ways to deal with time-varying parameters. One is to seek a "deep structural" model of money demand, i.e., one that is invariant to financial innovation and monetary policy regime changes. In order to achieve invariance, however, a deep structural model would have to incorporate decision rules

that govern financial innovation as well as rules that govern monetary policy regime changes. This approach is attractive in principle, since it would enable economists to evaluate the effects of policy changes. But this route does not seem promising at present, since monetary theory has not yet advanced to the point where it can deliver empirically useful representations for these decision rules. Given the state of knowledge, it may be worthwhile to seek an alternative solution.

Another approach is to learn to adapt to functional instability. There are at least two ways to think about adaptation. The most common approach is to respecify the model's functional form when it goes seriously off course. For example, recent efforts to respecify M2 demand models are described in Feinman and Porter (1992) and Duca (1993).[1] Another approach is to apply time-varying parameter estimators to conventional models in order to allow parameter estimates to adapt quickly when shifts occur. The time-varying parameter approach may prove useful for forecasting, even if its role in evaluating proposed changes in policy rules may be limited.

These two approaches should be regarded as complementary. Functional respecification is an ex post activity and therefore is not useful at the onset of a turbulent period. It generally takes many years to recognize model instability and to correct the problem, and the time-varying parameter approach may pay important dividends in the interim. Furthermore, when a model is respecified, it may be worthwhile to re-estimate by a technique that gives greater weight to recent data and less weight to older data, and this is precisely what time-varying parameter estimators do. On the other hand, if new financial instruments are introduced or if there are important omitted variables, time-varying parameter estimators may never fully adapt, and functional respecification may be necessary.

This paper concentrates on the potential usefulness of time-varying parameter estimators and does not explore functional respecification. My goal is to provide some insight into the marginal value of time-varying parameter estimators, but I do not claim that this is the only way to proceed.

II. Recursive Estimators

In real time, Federal Reserve economists reestimate money demand models as new data become available. Since I want to study the reestimation process, it is useful to pose the problem in terms of recursive estimators. Begin by writing the model as

$$y_t = x_t' b_t + u_t,$$

where x_t and b_t are kx1 vectors, y_t and u_t are scalars. The vector b_t denotes the OLS parameter estimate based on data available through date t. If the model is reestimated by OLS each period, then b_t evolves as

$$b_t = b_{t-1} + P_{t-1}x_t (y_t - x_t' b_{t-1}) / (1 + x_t' P_{t-1}x_t),$$

$$P_t = P_{t-1} - P_{t-1}x_t x_t' P_{t-1} / (1 + x_t' P_{t-1}x_t),$$

where $P_t = (X_t'X_t)^{-1}$ and $X_t = (x_1, \ldots, x_t)'$. This is simply the formula for recursive OLS.

Recursive OLS might be appropriate if b were time invariant. However, theory and experience do not support time invariance. We regularly experience parameter shifts in money demand models. After a shift, the model tracks real balances poorly for a while, until the OLS recursions catch up with the parameter shift. The problem with recursive OLS is that it takes too long to catch up. Thus it seems worthwhile to consider alternative estimators that catch up more quickly.

Recursive OLS gives the same weight to all observations in the sample. When the model is subject to parameter shifts, it may be more sensible to give more weight to recent observations and less weight to distant ones. This intuition can be formalized in terms of discounted least squares (DLS) (Harvey 1981). That is, choose the vector b which minimizes the discounted sum of squared errors:

$$DSS = \sum_{t=1}^{T} \delta^{T-t} (y_t - x_t'b)^2.$$

The parameter δ is a discount factor. If $\delta = 1$, each observation is given equal weight, and this simplifies to OLS. If $\delta < 1$, observations close to the end of the sample (i.e., those close to the present) get more weight than those in the distant past. If the model is reestimated period by period by DLS, the parameter vector evolves as follows:

$$\tilde{b}_t = \tilde{b}_{t-1} + \tilde{P}_{t-1}x_t(y_t - x_t'\tilde{b}_{t-1}) / (\delta + x_t'\tilde{P}_{t-1}x_t),$$

$$\tilde{P}_t = \delta^{-1}\tilde{P}_{t-1} - \delta^{-1}\tilde{P}_{t-1}x_t x_t'\tilde{P}_{t-1} / (\delta + x_t' P_{t-1}x_t),$$

where $\tilde{b}_2$ denotes the DLS estimate based on data available through period t and $\tilde{P}_t^{-1} = \Sigma_{j=1}^{t}\delta^{t-j}x_j x_j'$. When $\delta = 1$, these recursions simplify to recursive OLS. When $\delta < 1$, the most recent observation gets more weight in the updating formula than it does under OLS.[2]

1. Feinman and Porter investigate alternative measures of opportunity cost with an emphasis on modeling effects of the steep yield curve. Duca proposes that mutual funds be added to M2 in order to internalize portfolio substitutions.

2. This technique is similar to the random walk parameter model of Cooley and Prescott (1976). One advantage of the Cooley-Prescott

The rationale for using DLS is that it will adapt more quickly to a parameter shift than will recursive OLS. But this comes at the expense of a loss in precision. For example, if the parameters were time invariant, DLS would discount useful information contained in the early observations, and this would increase the variance of the estimates. The parameter δ controls the terms of the trade-off. A value close to 1 favors precision over adaptability. A value far from 1 allows the model to adapt quickly but may produce highly variable estimates even when no shift has occurred. The discount factor must be chosen to balance adaptability against precision.

III. EXPERIMENTAL DESIGN

Robert Lucas warns economists to "beware econometricians bearing free parameters," and the DLS algorithm has a free parameter. Thus it is important to impose some discipline on the choice of δ. In particular, δ must be chosen based on information that is available before the forecast period begins. This section explains how δ is chosen and how the DLS algorithm is evaluated.

I divide the sample into three subperiods. The first covers the period 1954 to 1980 and is used to generate initial parameter estimates. M2 was redefined in 1980, and one of the criteria was that the new aggregate have a stable demand function (see Judd and Trehan 1992). Since parameter instability is not a problem for this subperiod, initial estimates are computed by OLS.

The second subperiod covers 1981 to 1988, and it is used to determine an optimal value for δ. I experiment with values of δ ranging from .8 to .99 and choose the value that minimizes the mean square error of recursive DLS forecasts.[3] M2 demand functions were relatively stable during this period. By choosing δ to optimize goodness of fit over this period, we ensure that the DLS algorithm produces reasonably stable parameter estimates during stable times. This is an important criterion. An algorithm that produced unstable estimates during stable periods would be of no use to anyone.

Some data are saved at the end of the sample to test the DLS algorithm. For the period 1989 to 1992, recursive DLS estimates are computed using the discount factors determined above, and they are compared with recursive OLS estimates. The principal reason for choosing 1989 as the beginning of the test period is a desire to have several years of data available for evaluating the time-varying parameter forecasts, and the results are not sensitive to the precise choice of sample split. Conventional M2 demand functions were unstable over this period. If my intuition is correct, the DLS algorithm should adapt more quickly than OLS, and recursive DLS forecasts should therefore have lower mean square error than recursive OLS forecasts.

IV. RECURSIVE ESTIMATES OF MONEY DEMAND

Basic Specification

This section applies recursive OLS and DLS to a number of standard money demand models. I assume that all the relevant variables are integrated processes and that there is a stable long-run relation between real balances and the scale variable.[4] Given these assumptions, money demand can be expressed as an error correction model. I consider a number of simple specifications which differ according to their scale and opportunity cost variables. The general specification is as follows:

$$ln(m_t / p_t) = a_0 + a_1 ln(w_t) + z_t,$$

$$b_0(L)\Delta ln(m_t / p_t) = b_1 + b_2(L)s_t + b_3(L)\Delta ln(w_t) + b_4 z_{t-1} + u_t,$$

where s_t is a vector of interest rate spreads and z_t is the long-run "equilibrium error," in the language of Engle and Granger (1987). The first equation defines the long-run equilibrium relation between real balances and the scale variable.[5] Interest rate spreads are stationary and thus do not belong in the cointegrating relation. The second equation describes the short-term dynamics. The presence of the long-run "equilibrium error" in the second equation ensures that the short-run adjustments in money growth ultimately lead back to the long-run equilibrium level of real balances; hence the name "error correction" model.[6]

approach is that it implicitly allows different discount rates for different parameters. However, this would violate Lucas's dictum to avoid proliferating free parameters. When the Cooley-Prescott model is restricted so that there is only one discount factor, it is basically the same as DLS. I prefer DLS because it is more intuitive.

3. I also experimented with an a priori choice for δ, which was determined by the criterion that the discount function have a half life of 5 years. In general, this produced out-of-sample results that were superior to OLS but inferior to the data-determined value of δ.

4. The data do not contradict these assumptions.

5. I also examined models in which a_1 was set equal to one. This restriction implies that velocity is stationary and is equal to z_t. This had little effect on the result.

6. See Mehra (1991) for a more detailed exposition of error correction models of money demand.

One can write this as a single equation by substituting z_t from the first equation into the second.

I consider various combinations of scale and opportunity cost variables. The opportunity cost of holding M2 depends on the spread between returns on alternative assets and the own rate of return on M2. The latter is calculated by the Federal Reserve Bank of Richmond as a weighted average of the returns on the components of M2 (e.g., Mehra 1991). For alternative rates, I experiment with the six-month commercial paper rate and the 10-year Treasury bond rate.

For scale variables, I experiment with GDP and personal consumption expenditures. GDP is the standard scale variable in the money demand literature. Consumption can be motivated in two ways. First, consumption is the appropriate scale variable in cash-in-advance models (e.g., Lucas 1988). Second, as an empirical matter, various authors have emphasized that permanent income performs better than current income (e.g., Meltzer 1963), and consumption is a natural, observable proxy for permanent income.

To complete the specification, each version of the model also includes dummy variables for the second and third quarters of 1980, during which credit controls were binding, as well as a dummy variable for the first quarter of 1983, when MMDA accounts were introduced. Finally, the lag polynomials in the second equation are assumed to be of order 1. This is sufficient to capture the dynamics of real M2 growth during the initial estimation period.[7]

The Experimental Period

Each of these models was estimated by recursive OLS and DLS, using quarterly data, and the results are reported in Table 1. The first two columns report the mean square error of recursive one-quarter-ahead forecasts for the various models and time periods. Mean square error is standardized by dividing by the variance of the dependent variable; thus R^2 statistics are equal to 1 minus the mean square error. R^2 statistics are useful for evaluating absolute performance, and mean square error is useful for evaluating relative performance.

The first column of Table 1 reports results for the experimental period 1981–1988. During this period, the various models accounted for roughly 35 to 65 percent of the variation in real balance growth. GDP appears to be superior to consumption, reducing the mean square forecast error by roughly 20 percent. Furthermore, the short-term interest rate appears to perform better than the long-term rate. For example, when GDP is the scale variable, the mean square error for the short-rate model is approximately 35 percent lower than the mean square error for the long-rate model. Including long rates as well as the short rate has no effect on forecast performance. Thus, during the experimental period, the two best models were the ones based on GDP and short-term interest rates.

Even during this period, when M2 money demand was relatively stable, recursive DLS often worked significantly better than recursive OLS. In particular, for the models that include short rates, DLS reduces mean square error by an average of approximately 15 percent. In the long-rate models, the optimal discount factor turns out to be .99, so there is essentially no difference between DLS and OLS. Recall that DLS trades precision for adaptability. Even during the relatively stable subperiod, the gain from adaptability often outweighed the loss of precision.

TABLE 1

PREDICTIVE POWER AND BIAS

	MSE(81-88)	MSE(89-92)	BIAS(89-92)
1. Six-Month Commercial Paper Rate			
GDP			
$\delta = 1.00$	0.471	3.655	−5.42 (.000)
$\delta = 0.82$	0.358	1.430	
Consumption			
$\delta = 1.00$	0.601	1.564	−3.14 (.002)
$\delta = 0.80$	0.510	1.316	
2. Ten-Year Treasury Bond Yield			
GDP			
$\delta = 1.00$	0.637	1.267	−1.35 (.176)
$\delta = 0.99$	0.638	1.270	
Consumption			
$\delta = 1.00$	0.679	1.290	−1.42 (.154)
$\delta = 0.99$	0.681	1.269	
3. Six-Month Commercial Paper and Ten-Year Treasury Bond Rates			
GDP			
$\delta = 1.00$	0.482	3.768	−5.47 (.000)
$\delta = 0.80$	0.361	1.704	
Consumption			
$\delta = 1.00$	0.648	1.149	−2.71 (.002)
$\delta = 0.93$	0.613	1.109	

NOTE: The first two columns report the mean square error of recursive prediction errors scaled by the variance of real M2 growth. The third column reports the statistic $\sqrt{k}E(v_t)/\sigma$, with normal probability values shown in parentheses. When $\delta = 1$, this corresponds to recursive OLS.

7. Formally, this is sufficient to generate white noise residuals during this period.

The Test Period

The second and third columns of Table 1 report results for the test period 1989–1992. These columns reveal four results. First, when estimated by recursive OLS, the performance of all the models deteriorates badly. The mean square error of recursive OLS forecasts increases dramatically, and "out-of-sample" R^2 statistics are negative in every case (see the second column).[8]

Second, during the test period, the recursive OLS models consistently overestimated real M2 growth. One can test for bias in recursive OLS by computing the mean of normalized OLS prediction errors:

$$v_t = (y_t - x_t' b_{t-1}) / f_t^{1/2},$$

where $f_t = (1 + x_t' P_{t-1} x_t)$. If the model is stable, v_t has mean zero, is serially uncorrelated, and has the same variance as the regression disturbance. Further, if the regression disturbance is normally distributed, then v_t is also normally distributed (see Harvey 1981). Let $E(v_t)$ denote the mean of v_t over the test period: $E(v_t) = (1/k)\sum_{t=T+1}^{T+k} v_t$. Given these assumptions, $E(v_t)$ is normally distributed with mean zero and variance equal to σ^2/k. Thus $\sqrt{k}E(v_t)/\sigma$ is distributed as a standard normal random variable.[9]

Bias statistics are reported in the third column of Table 1. The mean recursive residual is negative in all models, and the means are statistically significant in four of the six cases. Since stable models have unbiased recursive residuals, this result confirms our belief that conventional money demand models should be treated as time-varying parameter models.[10]

Third, compared with OLS, DLS performs relatively well, and the percent improvement appears to be positively related to the degree of model instability. For example, the two models that had the lowest mean square error during the period 1981–1988 (i.e., the GDP-short-rate models) turn out to have the highest mean square error during the test period. For these models, DLS reduces the mean square error of one-quarter ahead forecasts by 55 to 60 percent. Thus DLS can be an important hedge against gross instability.

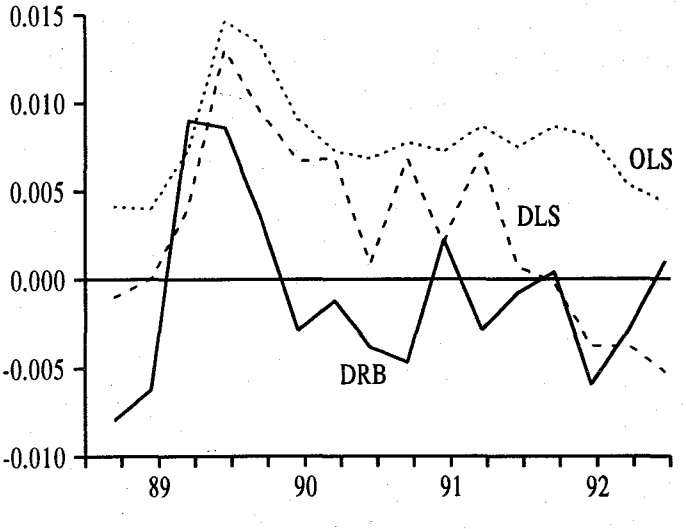
Figure 1 illustrates one of these cases. This is derived from the model that uses GDP as the scale variable and the six-month commercial paper rate as the opportunity cost variable (i.e., the first model in Table 1). The solid line shows real M2 growth, and the dotted lines show forecasts generated by recursive OLS and DLS, respectively. In the second half of 1989, both models systematically begin to overestimate real M2 growth. Recursive OLS is slow to catch on to the apparent structural shift, and it continues to overestimate real M2 growth until the end of 1992. Recursive DLS is more adaptable. It begins to recognize a structural shift around the second quarter of 1990, and its forecasts begin to edge downward. By the second half of 1991, this model seems to be back on track. Whether it stays on track remains to be seen. In this model, recursive DLS reduces mean square error by 60 percent relative to recursive OLS.

DLS is least useful in models that are relatively stable. For example, in models that omit the short-term interest rate, DLS and OLS produce basically the same results. It is worth noting DLS does not significantly *hurt* forecast performance when applied to relatively stable models, so discounting appears to be essentially costless.

Fourth, despite the relative improvement due to DLS, the absolute performance of the DLS models also deteriorated badly during the test period. These models also consistently overestimate real M2 growth and also have negative out-of-sample R^2 statistics. Thus, while DLS is

FIGURE 1

RECURSIVE FORECASTS



8. Recall that R^2 equals 1 minus mean square error. Recursive prediction errors do not necessarily have mean zero, so their mean square error can be larger than the variance of the dependent variable.

9. $E(v_t)$ is asymptotically normal even if the regression error is not normally distributed, provided that the regression error satisfies a mixing condition (e.g., White 1984). Thus, $\sqrt{k}E(v_t)/\sigma$ is approximately normal for reasonably large k.

10. This result contrasts with Mehra (1992), who fails to reject parameter stability in a similar model.

better than OLS, it does not appear to generate enough improvement to revive the use of M2 as an indicator of short-term fluctuations in nominal aggregate demand.

In retrospect, it is clear that completely naive, atheoretical forecasts would have worked about as well as any of these money demand models over the period 1989–1992. For example, forecasts based on a random walk model of real balance growth would have produced a mean square error of 1.387 over this period, which is comparable to the performance of these models. This simply highlights the difficulty of using historical relationships to forecast during turbulent periods.

V. CONCLUSION

Conventional money demand models often exhibit parameter instability, and this complicates the implementation of monetary policy. Applied macroeconomists might respond to this in two ways. They might seek time-invariant, deep structural representations, or they might apply time-varying parameter estimators to conventional representation in order to allow parameters to adapt quickly when shifts occur. This paper takes the latter approach, exploring the properties of recursive discounted least squares. This technique is designed to give greater weight to more recent data and less weight to older data, and this makes it more adaptable than recursive OLS.

The results suggest that DLS may have a useful but limited role to play in policy modeling. During unstable subperiods, DLS works better than OLS, and the gains can be substantial. For example, in a standard money demand model, in which the scale variable is GDP and the opportunity cost variable is the spread between commercial paper rates and the own return on M2, DLS reduces the mean square error of one-step-ahead forecasts by 60 percent. On the other hand, the absolute performance of DLS estimators also deteriorates badly over the last few years, and the models do not deliver reliable forecasts of M2 money demand. Thus it will continue to be difficult to interpret fluctuations in money growth.

REFERENCES

Cooley, T.F., and S.F. LeRoy. 1981. "Identification and Estimation of Money Demand." *American Economic Review* 71, pp. 825-844.

_____, and E.C. Prescott. 1976. "Estimation in the Presence of Stochastic Parameter Variation." *Econometrica* 44, pp. 167-184.

Duca, J.V. 1993. "Should Bond Funds Be Included in M2?" Working Paper No. 9321. Federal Reserve Bank of Dallas.

Engle, R.F., and C.W.J. Granger. 1987. "Cointegration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55, pp. 251-276.

Feinman, J.N., and R.D. Porter. 1992. "The Continuing Weakness in M2." Working Paper No. 209. Board of Governors of the Federal Reserve System.

Goldfeld, S.M. 1976. "The Case of the Missing Money." *Brookings Papers on Economic Activity* 3, pp. 683-730.

Harvey, A.H. 1981. *Time Series Models*. London: Philip Allan.

Ireland, P.N. 1992. "Endogenous Financial Innovation and the Demand for Money." Working Paper 92-3. Federal Reserve Bank of Richmond.

Judd, J., and J. Scadding. 1982. "The Search for a Stable Money Demand Function: A Survey of the Post-1973 Literature." *Journal of Economic Literature* 20, pp. 993-1023.

_____, and B. Trehan. 1992. "Money, Credit, and M2." Federal Reserve Bank of San Francisco *Weekly Letter* (September 4).

Lucas, R.E. 1988. "Money Demand in the United States: A Quantitative Review." *Carnegie-Rochester Series on Public Policy* 29, pp. 137-168.

Mehra, Y.P. 1991. "An Error Correction Model of U.S. M2 Demand." Federal Reserve Bank of Richmond *Economic Review* 77, pp. 3-12.

_____. 1992. "Has M2 Demand Become Unstable?" Federal Reserve Bank of Richmond *Economic Review* 78, pp. 27-35.

Meltzer, A.H. 1963. "The Demand for Money: The Evidence from the Time Series." *Journal of Political Economy* 71, pp. 219-246.

White, H. 1984. *Asymptotic Theory for Econometricians*. New York: Academic Press.

Water Policy in California and Israel

Ronald H. Schmidt
and Steven E. Plaut

The authors are, respectively, Senior Economist, Federal Reserve Bank of San Francisco, and Senior Lecturer in Business Administration, University of Haifa, Israel.

Water policies throughout the world often avoid market-determined allocations. In this article, we focus on case studies of Israel and California. Despite major cultural and political differences, it is found that water is heavily controlled through similar administrative mechanisms in both areas. Moreover, in both cases, these controls have led to inefficient allocation schemes favoring agriculture at the expense of other uses. This article examines the institutional factors that have led to such controls, and argues that adopting a new regulatory framework similar to that used to regulate electricity can still meet social concerns while dramatically improving economic efficiency.

Throughout history and in many parts of the world, water has been treated differently from other commodities. In fact, there are many that argue that water is not a commodity, but is special because of the role it plays in human survival and development. Dry areas that have been able to gain access to water have turned from deserts to gardens, while areas that have been deprived of water become wastelands.

Because of water's importance to dry regions, access to water is typically a matter of public policy. As suggested by the old saying in the West, "whiskey is for drinking; water is for fighting over," allocation of the resource tends to be the result of contentious historical experience. Development of water tends to be the outcome of publicly sponsored efforts aimed at achieving larger social aims, and allocation of that resource is closely monitored.

It is the heavy restrictions placed on the allocation of water that make the resource so unusual, particularly in market-driven economies. Other resources have been developed in partnership with the government, and the government often has a say in how that resource is used. But, in most cases, government is concerned with initial allocations, and allows subsequent trading to occur to achieve improved outcomes for the recipients. In the case of water, such trading is restricted to a much greater extent.

This unusual control over allocation and use is especially apparent when viewed across cultures. In this article we examine water allocation and use in two very different political and cultural systems—Israel and California—which, despite major differences in nearly all other social and economic institutions, have remarkably similar policies for allocating water. Furthermore, population pressures have increased sharply in both of these areas, placing increasing stress on available water supplies. Additional insights, therefore, are possible by viewing how the two regions are coping with the growing shortages.

While the focus is on these two very different regions, the lessons are more broadly applicable. For example, most parts of the western United States face similar challenges with water allocation. Details of water administration vary by region, however, putting an exhaustive evaluation of all approaches to water allocation beyond the scope of this paper. Moreover, although there are some differences in practice—for example, Colorado allows some trading in a formal market, while Arizona ties water rights to property

rights in some cases, making it possible to buy agricultural land and transfer the water to the cities—for most dry regions, the same basic problems dominate.

This article compares water policy in Israel and California to the U.S. allocation policy of electricity, another good characterized by increasing returns to scale. As discussed in section II, water shares many physical characteristics with electricity, particularly in the infrastructure and institutions needed to develop and get the product to consumers.

The major finding is that water policy differs from electricity policy in one key respect: After the initial allocation, water is much more controlled than electricity, with significant limitations on trading. Thus, it is allocated by quantity rationing. Electricity allocation, on the other hand, has some aspects of social allocation, but, ultimately, prices are used to ration demand (even if they are subsidized prices for some users).

Because of this administrative control, we argue that high efficiency costs are observed in the water delivery systems of the two regions—costs that are directly attributed to the restrictions on trading. In both regions, water is used in ways that force the development of inefficient, high-cost alternatives.

The purpose of such public micromanagement is apparently to control more fully the pattern of economic development in the region. Reform efforts in the two regions provide a useful contrast in this respect. In California, economic forces have become increasingly important in pushing for water allocation reform. Major efforts are underway to change the allocation mechanism to increase the role of market forces—that is, to allow trading. And one factor contributing to the success of those efforts is the declining relative importance of the state's primary water user—agriculture—and the growing political and economic power of cities and industries. In Israel, however, such reform efforts remain weak, and the government's need to control the use of water—for strategic and political purposes—continues to dominate the economic factors that are pushing for reform.

This article is organized as follows. Section I presents institutional details on current water allocation systems in the two regions, including both the physical structure of the water delivery systems and the economic and political infrastructures used to allocate the water. Section II compares those allocation mechanisms to the relatively more market-oriented mechanisms used to allocate electricity in the United States. We look at the reasons for government involvement in developing and allocating a resource like water, and we examine the extent to which the regulation of electric utilities provides a viable alternative model for water allocation. Section III discusses the costs of not using an electric utility-style allocation scheme, pointing out the inefficiencies resulting from the implicit ban on exchange. Section IV discusses traditional responses to rising shortages, and indicates the problems that have arisen in continuing "business as usual" in the two regions' water policies. Section V discusses some of the reform efforts underway in California, and Section VI presents conclusions.

I. Water Institutions and Policy in California and Israel

California and Israel share a number of similarities in water policy, despite other cultural and economic differences. Both have large semi-arid areas, mountain ranges, and mild, wet winters (except in extreme elevations) combined with dry rainless summers. Both regions experience rainfall that is concentrated in the north, and which then must be transported and pumped to drier southern regions. Both use extensive networks of aqueducts, pipelines, and pumping stations, and both have extensive experience with "high-tech" irrigation technology and biogenetic engineering in agriculture. Moreover, despite very different political cultures, both regions rely extensively on political and not economic policies for water allocation.

Israel

In an average year, Israel draws 1.2 to 1.3 million acre feet (MAF) of fresh water; 60 to 75 percent is consumed by agriculture (depending on supply conditions each year) and the residual goes to urban and industrial consumers. This water is obtained from several sources. The main body of fresh water in Israel is the Kinneret, or Sea of Galilee, which provides about half of the total supplies. It is located in the Jordan rift and sits well below sea level. This means that expensive pumping and transport is required to move water from the Sea to the country's farming areas, most of which are above sea level. The Sea is fed by the Jordan river and also empties into the same Jordan river, which then flows south into the Dead Sea. The latter is even further below sea level; indeed, it is the lowest point on earth and resembles somewhat California's Mono Lake. Water is conveyed from the Jordan River to the southern part of Israel through the National Water Carrier (a pipeline and aqueduct system), which was completed in 1964.

In addition to the Jordan river and the Sea of Galilee, there are two large underground aquifers, the Coastal and the Mountain aquifers, the latter encompassing central Israel and much of the West Bank. Water drawn from these sources accounts for the bulk of the remaining supplies. A variety of wells, oases, and dams capture water from

additional sources, but provide relatively small quantities.

Israel also makes considerable use of recycled waste water, particularly for irrigation purposes, as well as some brackish water from wells near the seacoast and near Eilat. Desalination of sea water has been used only on a limited or experimental basis.

Proposals for new sources of water that have been considered include large-scale desalination, water imports from Turkey, and even a canal from the Mediterranean to the Dead Sea, that would allow complete diversion of the Jordan river for commercial purposes. Other possibilities include new ground water discoveries, and diversions from the Litani River (Wolf and Ross 1992). In all of these cases, however, the projects may not be commercially viable and there are fears that they could cause extensive environmental damage.[1]

Under Israeli law, water is a nationalized public good. As specified in the 1959 Water Law, all water is the property of the state, including waste, sewer, and runoff water that can be commercially used. An owner of land does not own water under the land, and there are no riparian rights.[2] This legal status of water continues practices incorporated into the Fundamental Law defining the rules of government in 1949, and those embedded in the British Mandatory laws. The 1959 law essentially perpetuated the then-existing water allocation pattern, with water set aside for planned future settlements and activities. Water is supplied by Mekorot, Ltd., a public corporation that pumps and supplies about 60 percent of the nation's water, and by small private suppliers.

The Minister of Agriculture is the supreme statutory authority charged with formulation of water policy—including consumption, pricing, and allocation—subject to oversight by a Knesset water committee that must approve water pricing changes. The Minister appoints the Water Commissioner and an advisory Water Commission, and also the directors of other public sector agencies that play a role in water development, pricing, and supply.[3]

Some analysts have argued that this arrangement has fostered an automatic conflict of interest in water allocation. Because the natural tendency of the Ministry of Agriculture is to work as an advocate on behalf of farm interests, water policies also tend to be formed by agricultural interest groups (Galnoor 1978). In part because of this tendency, water policy in recent drought years has allowed substantial over-consumption of water supplies by agriculture, even though these actions have polluted and damaged the structure of the underground aquifers.

Even when cuts are made in supplies, the policies appear to be crafted with farm welfare in mind. For example, in 1991, the Minister of Agriculture implemented across-the-board cuts of 25 percent in agricultural water allocations since the Sea of Galilee's surface level had dropped below its "red line," a somewhat arbitrary level selected as the minimum allowable level.[4] Farmers whose allocations were cut were eligible for compensation from taxpayers for lost revenues that would have been generated with the water.

In Israel, water allocations tend to be political. Historical allocation is one guiding principle, with water users generally able to receive the same allotment in future years if they use the supplies they are granted in the current period. Apportionment of additional water often takes place subject to political pressure.

Once granted, water allocations in Israel are extremely inflexible. Farmers are allocated water to grow specific crops. If a farmer wants to change his crop mix, he must apply to the Ministry for permission to apply the water to that different crop. Allotted water not used cannot be sold—it is explicitly illegal to sell water or water rights in Israel. Violators are subject to criminal prosecution. Moreover, farmers who temporarily consume less than their full allotments may find their future allotments cut in subsequent years, creating the incentive for farmers to use all of their allocation of water to preserve future deliveries, even if the use is wasteful. Finally, a farmer who sells his land cannot sell his water allotment with the land, and must include a clause in the contract where the buyer attests to having been forewarned of this.

California

Water allocation in California is similar to that in Israel. On average, 80 to 85 percent of net water consumption occurs in the agricultural sector. Urban users consume 10 percent, with the residual allocated to industry.

California has one of the most intricate water supply systems in the world. Most of the rainfall and snow

1. According to a feasibility study by the Israeli government in 1983, additional use of water from the Dead Sea could cause minerals to cake on the surface. (The Dead Sea has salt concentrations 55 times that of the Pacific Ocean.) In addition to environmental damage, they concluded that such efforts would damage the tourist trade and mineral extraction processes in the vicinity.

2. Riparian rights are based on English law, and grant a landowner the right to use water that passes through his or her property. In most Middle East countries, mineral and water rights under one's land belong to the state.

3. At the present time, the powers of the water committee are being transferred to the Finance Committee of the Knesset.

4. Despite its designation as a minimum, the Ministry approved even further pumping below the "red line" in 1990 and 1991 because of low rain levels.

accumulate in the northern and eastern parts of the state, while most of the population is in the western and southern, semi-arid regions. A series of dams and reservoirs capture and store water in the Sierra Nevada and the northern part of the state for transport in a vast system of canals and aqueducts to the populated coastal and central agricultural regions.

California's water is developed and supplied by a variety of different agents. The two largest projects, the Central Valley Project (CVP) and the State Water Project (SWP), provide 27.5 percent of the net water supplied in the state in an average year. The CVP was developed and is operated by the federal government, while the SWP was developed and is operated by the state. The two systems provided 7.0 and 2.4 MAF of water, respectively, in 1985, the last year of relatively normal supplies.

In addition to those projects, individual cities have developed reservoirs and delivery systems, such as the Hetch Hetchy reservoir for San Francisco and the Lake Cachuma and Gibralter reservoirs for Santa Barbara. Los Angeles also has aggressively redirected water from other sources. Los Angeles receives water from the Owens Valley and Mono Lake through the Los Angeles Aqueduct, as well as some water from the Colorado River. Recent court decisions, however, have reduced future deliveries from these sources. Total withdrawals from the Colorado River were around 5 MAF in 1985; that total is expected to drop to 4.2 MAF by 2010.

Finally, complementing these surface water sources is an extensive supply of ground water in aquifers. In 1985, sustainable ground water supplies were estimated to be around 6 MAF of the state's total of 32.2 MAF. An additional 2 MAF was overdrafted in that year, to yield total supplies of 34.2 MAF. During drought years, ground water is drawn more heavily, smoothing supplies from year to year.

Several new sources of supply are under consideration. According to the California Department of Water Resources (Department of Water Resources 1987), an additional 1.4 MAF will be needed by the year 2010 to meet existing and projected needs of an additional 6.5 million people. Most of that gain depends on further development of the state and federal water systems, including the Kern water bank, construction of the Auburn dam, and completion of Los Banos Grandes reservoir. Other potential sources include further conservation efforts, development of waste water re-use, and desalination plants.

California's water resources are administered by a large number of overlapping state and federal agencies. CVP water is federally administered by the U.S. Bureau of Reclamation, with water delivered to CVP contractors. SWP water is administered by the state. The state's Water Resources Control Board is the agency most directly involved in determining possible shifts of water from one user to another, but the Department of Fish and Wildlife and the federal Environmental Protection Agency, among others, also have critical input into the process. At the local level, water districts have the power to reallocate water within a district and often have veto power over shipments out of the district.

Rights are an accumulation of historical precedents. Riparian rights, establishing the right to use water that passes through one's land, apply to many of the water resources claimed early in California's development. Appropriative rights apply to most of the state's water, although the rules governing those rights differ depending on the date granted. Appropriative rights allow the user to divert water for "beneficial use," with rights sequentially based on when the right was granted. These rights were designed to protect early developers located downstream from losing water because of newer upstream diversions.

As in Israel, however, rights do not include automatic ownership. Water is deemed a public good owned by the people of the state. The "Public Trust Doctrine" is frequently cited by the courts in water disputes. With roots in Roman law, the doctrine of public trust holds that certain resources are the property of all. In a 1983 decision (National Audubon Society vs. Superior Court), the California Supreme Court held that the state has a duty to consider public trust values before it approves water rights applications or adjustments. This doctrine has been used most recently to guarantee water for in-stream, environmental uses.

Appropriative rights allow users to apply the water for beneficial purposes, but do not allow the rights holders to treat the resource as an asset. Thus, most water supplies cannot be sold or traded to other users without explicit approval of a variety of agencies, including the local water district, the Bureau of Reclamation, the state Water Resource Control Board, and possibly the EPA, U.S. Department of Fish and Game, and the state's Department of Fish and Wildlife. Moreover, water that is not used by a rights holder may be interpreted as surplus water that is not beneficially used. Like Israel, therefore, California's incentives are structured to "use it or lose it," with users that use less than their full allotment potentially losing that surplus in future years.[5]

5. Recent examples of this interpretation have emerged in California. Some rice farmers that idled fields and sold their water to the state's Emergency Water Bank in 1991 are reporting efforts by the state to reduce their allocation on these grounds.

II. IS WATER SPECIAL?

As discussed above, water allocation is highly controlled in Israel and California. Water is allocated according to historical precedent, with modifications to reflect changes in available supplies or competing uses made by administrative fiat. In drought years, for example, some users are given only partial allotments, with cutbacks either across the board or applied sequentially across classes of users.

Economic forces are largely ignored. Water transfers are difficult, even when they are arranged to the mutual benefit of both parties. Prices do not change to reflect growing scarcity, and hence, prices are not used as a tool to encourage conservation or as a mechanism to evaluate new infrastructures.[6] Differences in prices across regions and users are not used as signals to encourage transfers of water from low-priced to high-priced regions. In fact, those facing limited supplies are encouraged to develop new (usually higher cost) sources, rather than to purchase supplies from existing users, even when both parties could gain from the exchange.[7]

This treatment of water, which differs significantly from the way other resources are allocated in market-oriented economies, is typically justified on the grounds that water is special. In this section, we examine two related issues. First, we discuss the reasons that government typically intervenes in water delivery, drawing parallels with other regulated natural monopolies—particularly electricity. Second, we examine the differences between the allocation of water and of electricity and identify the root difference between regulation of those industries.

Arguments for a Governmental Role in Water Delivery

Several arguments are put forth in favor of involving government in the allocation of water in dry regions. Three issues typically are cited: the cost structure of water investments, other noneconomic public policy goals, such as encouraging migration and directing land use decisions, and concerns about equity and economic disruptions ("third party effects"). As we argue, however, each of these arguments also could be—and is—applied to electricity allocation, and the more successful experience of that industry in adjusting to growing and competing needs appears to offer a useful guideline to improving the current system.

Water Investment Costs. Public investment in new water facilities often is justified on the grounds of increasing returns to scale. Typically, the infrastructures needed to store and deliver new water supplies—dams, pumps, and canals—are governed by decreasing marginal cost structures. Slightly increasing the size of a dam or a canal can cause a large increase in capabilities since, in general, volume does not increase linearly with increased investment, but roughly geometrically. Consequently, marginal costs tend to fall with increasing project size up to some point.

Often the point of minimum marginal cost requires an investment too large for an individual or group of individuals to coordinate.[8] In those cases, the government often is asked to step in on behalf of its constituents.[9] For example, after considerable lobbying, the California Legislature authorized the Central Valley Project as a state water project in 1933, which the federal government eventually built and operated. The other major water projects in California also are the result of state or local government efforts, including the SWP.

---

6. In fact, prices are used in what appears to be a punitive fashion by many water districts. During the height of the drought in California, for example, moral suasion (rather than price increases) and threats of restricted service caused urban users to reduce consumption sharply. Water districts then were faced with revenue shortfalls, since prices had not adjusted. They then had to raise rates to consumers. Thus, water districts were in the awkward position of penalizing consumers for doing what the water districts had requested. Obviously, if the districts had instead used higher prices to discourage water consumption, this problem would have been avoided.

7. Interestingly, while lack of transfers often forces urban areas to consider more exotic technologies, pricing policies make those investments appear unreasonable. For example, Israel recently decided against a major water resource development project in the southern port city of Eilat because it was not viable at current water prices. However, since prices often are based on pumping and transportation costs, and ignore the social opportunity costs and scarcity value of the resource, those prices may understate the value of a new facility. Consequently, this project might have been viable with proper pricing.

8. According to Reisner and Bates (1990), the federal government tried to encourage private water development in the nineteenth century by offering free or highly subsidized land to those that would undertake such development. However, because of the large costs involved, private efforts were largely unsuccessful. "At the eighth National Irrigation Congress in 1898, one speaker compared the western landscape to a graveyard, littered by the 'crushed and mangled skeletons of defunct irrigation companies . . . which suddenly disappeared at the end of brief careers, leaving only a few defaulted obligations to indicate the route by which they departed' " (p. 13).

9. Work by Buchanan and Tullock (1962) and Olson (1982) has argued that the process by which projects such as these are developed can be explained by special interest group models. Concentrated single-issue interest groups often are able to obtain political support and sufficient votes to pass legislation favorable to that group because the costs imposed on the non-beneficiaries are spread thin, making political opposition weak. The beneficiaries of new water projects typically are relatively concentrated. In California, for example, the CVP delivers

In Israel, the government also has been the principal agent of water development. Galnoor (1978, p.343) characterized the rationale for government involvement as follows:

> Divergences between the private side and the public side of water costs and/or benefits, as well as the need for high initial investments and the characteristics of a natural monopoly, contribute to the necessity of some form of public intervention in the management of water resources. In Israel, government intervention is also required because water is a part of the infrastructure for (a) the ideology of nation building based on farming and new settlements and (b) economic development.

The Israeli government, therefore, has taken upon itself the responsibility for planning, constructing, and maintaining new facilities. Through 1970, gross investment in water projects by the government was estimated to be equal to between 3 and 5 percent of total gross capital investment in the country. Costs of new supplies were so high that expansion plans involving the United States as a partner in a joint venture to build large desalination plants were developed in the mid-1960s, although the plans were never implemented (Tahal 1972).

Development Tool. A second motivation for government control over water allocation is water's power as a tool to influence migration and land use. In Israel and the western United States, pro-growth forces were very strong during the periods of water development. Often one of the most powerful inducements to potential migrants was access to cheap land in peripheral areas. In much of California and Israel, the available peripheral land was not particularly attractive to settlers unless water could be provided to irrigate crops. Thus, in both regions, policies often were designed to create large supplies of dependable water for agricultural purposes.[10]

In Israel, strong priority was given to encouraging immigration and population dispersal. Development of available agricultural land has been an important goal (Galnoor 1987, p. 345):

> In the context of Zionist ideological objectives, water has never been regarded as merely another economic resource, but as a prerequisite to efforts to create a new society in the (cultivated and redeveloped) Land of Israel. The selection of one water project over another was not determined on the basis of relative economic returns. This ideology stressed a "productive occupational pyramid" based mainly on farming in the collective Kibbutzim and in the smallholders' communal settlements.

As discussed by Plaut (1992, p. 16), Israeli policy also has sought to encourage active cultivation and occupation of its lands for national security reasons:

> There is an ancient and broad consensus in Israel that survival of the state requires "settlement" of the land by Jews. In many cases, "settlement" is taken to mean farming. The origins of the doctrine go back to the early phases of the Zionist movement, when the boundaries of "settlement" were believed to establish the geographic-political blueprint for a later state.
>
> Once Israel came into being, its borders were determined by either diplomacy or force of arms, but no longer by farm settlement. Nevertheless, the belief that land settlement provides political and strategic control of territory has survived. It is closely linked with the ideological consensus in favor of population dispersal policies.
>
> Initially, dispersal of farm settlements along frontiers played a strategic role, making border patrol and surveillance easier. Later, it was argued that these settlements could provide support and services for army units stationed near the frontiers. After 1967, agricultural settlements were established in the Territories for the express purpose of creating political "facts" and new strategic realities. In any case, it is widely believed that an absence of Jewish settlement in any part of the country may lead to loss of that region through international pressure and/or Arab insurgency.

More recently in California, water policy has been used in the reverse direction—preventing growth. Barriers are raised to converting agricultural water use to urban use, in part to prevent conversion of agricultural land to urban and industrial purposes. Opposition to water markets in California's Central Valley, for example, is driven by fears that water markets would encourage the transition away from agriculture and bring major changes in the industrial and social fabric of the valley communities. In other cases, such as in Santa Barbara, the city council chose in the early 1970s not to hook up to the State Water Project, arguing that access to additional sources of water would allow more

---

water primarily to agricultural users. Similarly, in Israel, farmers consume three-fourths of the total supply. The water is subsidized by other sectors of the economy, which bear the residual cost of operating and constructing supply facilities. Additional federal projects continue to be proposed and constructed, where the costs are borne nationally and the benefits are concentrated in specific regions and groups of consumers. Thus, according to this theory, projects can be approved that provide subsidized water for some users that would be prohibitively expensive to the beneficiaries if they were developed privately by those beneficiaries.

10. According to Reisner and Bates (1990), part of Theodore Roosevelt's motivation for developing water projects in the west was to build up "America's weak western flank" (p. 14).

people to move to the area and create excessive growth.[11]

Fairness. A third argument for government allocation of water is concern over fairness and the potential costs of changing existing allocations. Fairness issues arise whenever reforms are considered. Since current policies explicitly state that water is a public good, held in trust by the State and is not owned by any individual, any change whereby an individual gained legal title to the water would involve a change in the distribution of wealth and income. In such cases, it becomes possible to ask whether current subsidized water recipients "should be" entitled to those resources, or whether another allocation is more fair.

Fairness also appears in the debate because of fears that pricing the resource would make water unavailable or too costly to low-income persons and farmers. Some interest groups voice concerns that market prices would be too high for many consumers, so only the wealthy would be able to afford the resource.

More generally, pricing policies of many projects explicitly recognize other social goals relating to fairness. For example, prices of CVP water are heavily subsidized. Most contracts called for fees that often were designed to pay the nominal cost of the construction over time. While initially designed as a 10-year repayment period, most Bureau of Reclamation projects eventually extended terms for as much as 80 years. Moreover, interest costs, under the Reclamation Act of 1902, were not charged (Reisner and Bates 1990).

In Israel, water "doctrine" has always been based on pricing formulas that reflect the farmer's "ability to pay" and not the scarcity value of water as a resource. Under this doctrine, a drought that reduced a farmer's ability to pay should produce lower water prices, not higher prices to farmers, whereas in a market system the price would rise due to the reduction in supply. Hence, "fairness" to farmers, rather than efficient use of the water, appears to be the predominant sentiment guiding water policy in Israel.

A related issue involves potential disruptions that any change in allocation might create, or "third party effects." Policymakers predict that the use of water markets would decimate agriculture and the agricultural communities by encouraging farmers to sell all of their supplies (causing large third party effects on other rural businesses). Large differentials in prices between cities and agricultural water districts are taken as evidence that markets would lead to large diversions between farms and cities, with water costs pricing agriculture out of water. And experience in Owens Valley, where Los Angeles acquired rights to water from the valley and transferred it to the city at the expense of the local economy, is frequently cited as a warning for the potentially negative effects of allowing transfers.

Water vs. Electricity

While these arguments have been powerful justifications for current allocation policies—and may argue for some governmental involvement in the system—they are not unique to water. In fact, these same arguments can be applied to electricity and other utilities, and yet those industries appear to be more adaptable and efficient than the water delivery system. In many instances, water appears to share more similarities than dissimilarities with electricity; the major exception is the way it is controlled after the initial allocation.

Natural Monopoly. Water and electricity have similar physical and technical characteristics. Indeed, they often are joint products of large dams constructed along rivers. Water and electricity both flow in complicated grids over long distances, and are delivered to municipal customers through a centralized utility. Technical characteristics are similar: Storage (dams and reservoirs), transportation (canals and pipelines), and distribution networks all exhibit economies of scale that lend themselves to the creation of natural monopolies. Marginal costs tend to fall over a large range, often making it inefficient to promote competition in many parts of the system.

But in the case of electricity, considerable work has gone into designing regulations that maximize the efficiency of the utility while recognizing these scale economies. Utilities that build generating plants are allowed to add costs of approved facilities into the rate base, and owners of the utilities are granted a rate of return on that capital.

Nonetheless, while highly regulated, electricity demand is rationed by price—except in the rare instance of power failures and shortages. Users are charged in ways to generate the necessary rate of return, cost of maintaining the facilities, and costs of inputs. Pricing schemes typically rely on average cost of delivery. Moreover, reforms are constantly being evaluated. For example, time-of-day pricing has been tested in several sites to match marginal costs more closely to prices charged for the resource.[12]

11. This policy has changed as a result of the recent drought. Extreme reductions imposed on homeowners led to political pressure to add new supplies. Santa Barbara did connect a small pipeline to southern California in 1991 to purchase water from the Metropolitan Water District, and it is exploring desalination options.

12. The key advantage of using prices with subsidies to ration demand, as opposed to direct allocation, is that in a pricing environment trading occurs. Even though a system with subsidies results in a distorted use pattern, it is still the case that recipients of the resource balance their valuation of the water against that of all other potential users. In direct quantity rationing without trading, mutually welfare-enhancing improvements from the initial distribution are not allowed.

Major reforms also have been suggested (and elements tested) to introduce more market forces into electricity distribution. As reviewed by Schmidt (1987), increasing interlinkages between power grids have made bulk power sales among utilities at market prices a least-cost mechanism to avoid unnecessary construction of costly new facilities. Other parts of the system also have the potential for injecting additional market forces to improve the operating efficiency of the system.

In many ways, the regulatory structure used for electricity appears applicable to allocating water. A regulated monopoly could be granted to the producers, transmitters, and distributors of the water. In principle, rate of return compensation could be arranged (paid to the government in the case of state or federal projects), with the price to consumers ultimately serving to ration supplies among consumers. Moreover, like bulk power, water could be sold among primary owners of water at market prices, and moved (analogous to "wheeling" in the electric utility industry) along the canal network to its final destination.

Development Tool. Water allocation often is intended to guide economic development. Prices are established to subsidize and encourage use in particular areas. The power of water has been readily apparent in both California and Israel.

Electricity has been used for similar purposes. The Tennessee Valley Authority was created in the United States to generate low-cost power to a large underdeveloped rural area. Similarly, the Rural Electrification program has the objective of bringing low-cost power to rural areas to help speed development in those areas.

Electricity, like water, is viewed by many governments as one of the basic infrastructure ingredients necessary to promote economic growth. Encouraging access to electricity networks has been a central part of many economic development programs. Thus, the electricity model of allocation appears rich enough to encompass these additional goals of water development.

Fairness. Concerns about social equity often work through water pricing policies. Prices for water generally are tied to historical construction costs—costs that are typically well below the economic value of the resource.[13] In Israel, even this tenuous link to construction costs is missing. The Water Law requires that "in spite of differential costs, water prices in the various regions be equalized. In practice, water charges have been relatively uniform and quite often nominal" (Sadan and Ben-Zvi 1987, p. 3).

In electricity allocation, similar goals exist and are accommodated through pricing policies. The same concerns often are voiced about low-income consumers. In many states, electric utilities are prohibited from cutting off service to low-income consumers during the winter if they are unable to pay. Similarly, lifeline rates are offered to low-income, elderly, and handicapped individuals to assure their access to the resource. Finally, prices charged to industrial, commercial, and residential consumers are allowed to be structured in different ways to encourage certain uses.

In periods of temporary shortages, electricity policy also is designed to recognize social objectives. Since prices cannot be instantly adjusted and communicated to users, temporary surges in demand are met by graduated cutbacks to particular users. Heavy industrial users are cut back first, with critical needs (for example, hospitals) the last to be curtailed. Contingency plans for "brown-outs" and similar emergencies are established by utilities and approved by regulators to be consistent with social policy.

III. Implications of Trading

Water policy and electricity policy, therefore, share many of the same objectives and characteristics. Both often rely on government investment policies, seek to direct the pattern of economic development, and seek to redress social inequities by designing pricing and access policies to protect certain interests.

The key institutional difference between water and electricity results from the assignment of ownership rights. In the case of electricity, the units are clearly owned by some entity, and that entity has the right to distribute units to any customer or other utility. Even in the case of a publicly constructed facility, such as the Tennessee Valley Authority or Bonneville Power Administration, a public entity has clear ownership of the electricity, which it sells to utilities—sometimes subsidized, and sometimes according to allocation formulas. But there is no requirement that the receiving party must use that electricity, and thereby prevent the utility from transferring the power elsewhere.

In the case of water, transferability is severely restricted. As noted earlier, allocations are determined administratively, for the most part, and those allocations then are fixed. Contractors must use the water, or lose the rights to it.

Consider how such a policy would work in the electricity industry. Customers would receive a given supply of electricity at a particular time. If the customers did not use

13. A report by the Western Governor's Association (1987, p. iii) concluded: "The structure of the West's water system at federal, state, and local levels was designed to promote economic development through assuring a secure supply of water and to protect property rights in water once they were established. Laws, policies, and practices are largely silent on increasing efficiency of use."

that electricity, they would forfeit future rights to that resource. Such a system would lead to inefficient optimizing behavior on the part of consumers similar to that observed in the former Soviet bloc. In Poland, for example, heat was supplied at very low cost at specified times. The incentives built into the system, therefore, led to a practice of regulating temperatures by opening windows, rather than adjusting heat consumption. In other words, the incentives will lead consumers to use all of the subsidized electricity made available, since the alternative is to lose access to the power in the future.

The regulation applied to electric utilities demonstrates the potential to separate efficiency concerns—how the resource is used—from equity concerns. This separability is well-developed in the economics literature. This latter point is attributed to Coase (1960), who examined the importance of initial endowments in determining the final consumption distribution of a given resource. He demonstrated that if trading were allowed and transactions costs were small, the final allocation of a resource would be efficient regardless of the initial distribution of rights. This efficiency would be achieved through trading among potential consumers until the resource was finally used in its highest valued uses.

The "Coase Theorem" predicts that if trading were allowed, the assignment of ownership rights to water would have little effect on how it ultimately would be used (although that outcome could be considerably different from the current mix of production resulting from inefficient allocation of the resource). Whether farmers were granted ownership and allowed to sell to cities, or vice versa, the ultimate outcome in water use would be approximately the same. Clearly, wealth would be distributed differently under the two cases, but the Coase theorem argues that such ownership assignment only affects how the final basket of goods is distributed among consumers, not what or how much is in the basket.[14] Thus, it is possible to allocate rights in the interest of boosting equity, with the recognition that trading will promote efficient use of the resource.

In the case of electricity, initial allocations often are granted with redistribution as a goal, and limits are placed on trading to ensure that those goals are not circumvented. For instance, utilities can sell trade surplus power, but they must first satisfy local demand. However, utilities also have incentives to make local demand more efficient by subsidizing insulation and energy conservation efforts.

In the case of water, similar limitations could be put in place. But if the water districts had ownership rights to the water, they would have more incentives to encourage water conservation to make additional water available to sell outside the district.

The principal benefit of assigning ownership rights to water is to permit exchange. After choosing an allocation scheme that satisfies desires for fairness, individuals can collectively be made better off by allowing them to engage in mutually beneficial trade. Even with regulatory restrictions in place to favor particular uses, the ability to trade encourages all parties to recognize the opportunity cost embedded in any given use of the water.

The cost of not allowing trading is well-documented (Reisner and Bates 1990, Schmidt and Cannon 1991). Agricultural water is heavily subsidized in California, with the price of water to urban users on the order of 10 to 20 times that of most agricultural users—even accounting for differences in transportation costs and processing facilities.

These apparent inefficiencies are illustrated by use patterns in both regions. In Israel, cotton uses a major portion of the country's water supply. Yet, according to some analysts, cotton generates negative value-added in Israel, with the implicit subsidies granted to the sector exceeding the revenues from selling the crop. In California, 40 percent of the state's water is used to grow rice, alfalfa, cotton, and pasture, even though these crops altogether account for only 0.2 percent of total state income.

Moreover, lack of transferability has made it necessary for cities to plan construction of desalination plants. Such plants would yield water at a cost in excess of $2,000 per acre foot at the same time that water used for some low-value crops is priced at $8. Since costs of transportation are on the order of $100 per acre foot between many potential transfer sites, this price differential suggests that both parties could be made better off by trading. The cities could forgo constructing expensive new facilities, while the agricultural sector would be encouraged to increase its efficiency in water use to free up the resource to sell to the cities.

Potential gains from trading water have been demonstrated in several recent cases. A classic example of the gains to be made from trading is the 1988 agreement between the Metropolitan Water District of Southern California (MWD) and the Imperial Irrigation District (IID). In that arrangement, MWD agreed to pay the cost of lining irrigation ditches in the Imperial Valley in exchange for the right to buy the saved water.

14. Some researchers have disputed this claim, arguing that different initial assignments of rights would lead to different final consumption bundles. In particular, if there is a difference between the amount a farmer would pay to get a unit of water ("willingness to pay") and the amount she would accept to sell a unit to another person ("willingness to accept compensation"), the outcome would depend on whether she had the initial right to the resource or whether she had to purchase that right from another.

Interestingly, this case provides strong evidence of the magnitude of inefficiency that resides in the current administrative system. Both parties were made better off by the transaction. Indeed, it was sufficiently in MWD's interest to make the trade that they were willing to pay for the infrastructure improvements as well as pay for the water. Clearly, IID also gained in the process, since they were faced with no additional cost, yet gained a windfall profit from selling water that would otherwise have been lost to the district. But the incentives in the current system—including uncertainty about future rights to water—prevented this transaction from occurring automatically.[15]

Similarly, in the recent drought, California experimented with an Emergency Water Bank, where water districts could sell water to a state body, which then could resell the water to other districts with shortages. According to work by Howitt (1991), the effect of the Water Bank was to idle some acreage of rice and pasture, while permanent crops and high-value crops continued to receive sufficient water. Such transfers were voluntary, and resulted in farmers receiving compensation automatically from urban water districts.

In sum, the key difference between allocation schemes for water and electricity—and the cause of the high level of inefficiency in water use—is the result of failure to assign ownership rights to water users, and hence, to allow trading of those rights. Granting water rights holders the ability to engage in mutually beneficial trades would put in place incentives to increase efficiency in use.

IV. Rising Shortages: Failures of Traditional Remedies

Drought conditions in the 1980s and early 1990s revealed serious deficiencies in the water delivery systems of both Israel and California. Both regions experienced extended droughts that stressed the available supplies beyond normal experience. Large cutbacks in water deliveries to farmers were required, and extensive conservation and rationing schemes were imposed on urban and industrial users.

In both cases rains finally arrived and eased short-term conditions, but the public in both Israel and California have become increasingly sensitized to the inefficiencies and costs associated with centrally planned and allocated water. Water policy in both regions, therefore, is facing intense public scrutiny, with the public less willing to leave decisions to the "experts." Moreover, the droughts also highlighted the increasing scarcity of existing supplies over the long term, with growing populations likely to make responses to future droughts even more difficult and costly. Both Israel and California expect growing populations, while few new sources of supply are scheduled to come on line without further investments.

Traditionally, the response to shortages has been to locate and develop new supplies. California's Department of Water Resources, for example, projects the need for 1.4 MAF of new water by the year 2010, based on the assumption that the needs of a growing population are met by new sources, rather than through reallocation of existing supplies. Estimates in Israel in the mid-1970s called for a shortfall of about 0.25 MAF developing by 1985, again to accommodate new uses, not to reallocate existing supplies. Projections of population growth imply that by the early twenty-first century water may suffice for urban users only, with no agricultural water in Israel.

The traditional response—to meet growing demand by adding capacity—has faced resistance in recent years, however. Three related reasons have combined to make new facilities increasingly difficult to undertake.

First, the environmental movements in California, and to a lesser extent in Israel, have challenged additional water development by focusing attention on the previously ignored environmental consequences of water projects. In California, current water use patterns have caused widespread damage to the San Francisco Bay Estuary (the Delta). Agricultural runoff has degraded water quality by increasing the nitrogen content of the water. Low flow conditions caused by excessive pumping of water through the Delta to the southern portion of the state have caused periods of reverse flow, where salt water is pulled into the Delta. Moreover, periods of low flow have raised water temperatures in the rivers, a development that has been linked to a sharp decline in the number of salmon that spawn in the Sacramento River. Finally, past development policies have reduced wetlands areas, destroying the habitat of a wide variety of fish and migratory wildlife.

Changes in operating practices in the Delta, which are likely to be mandated by the EPA and the federal Department of Fish and Game, may have a profound impact on California's water supply. Currently, over half of the state's fresh water passes through the Delta. Improved environmental quality is likely to result in reduced shipments of

15. For a description of the events in this case, see Reisner and Bates (1990). While heralded as a success by those advocating increased use of voluntary transfers, this agreement also highlighted the problems with the current system of regulation. Rather than happening voluntarily in response to recognition of the mutually beneficial trade possibilities, the agreement was triggered by a legal challenge by one of the farmers and an order by the State Water Resources Control Board. The process took over eight years to complete.

water to the southern portion of the state.[16] As much as 1.1 MAF of the state's water may be removed from consumption and applied for environmental purposes.[17]

In Israel the main ecological policy issue has been the destruction of its aquifers. Overconsumption of water from those underground resources has begun to damage the geological structures of the aquifers, threatening to destroy future water supplies. Moreover, additional use of water from the Dead Sea could cause further environmental damage to that body of water.

Second, new supplies are expensive. The relatively inexpensive projects have already come on line. Currently, planners are considering new dams in California (although with little likelihood of success given environmental opposition), water banking in ground water basins, and the development of some new facilities to store water south of the Delta. But most plans for new supplies also involve desalination plants that can cost in excess of $2,000 an acre foot, and waste water treatment facilities that also yield high-cost water. Other choices involve forced rationing in urban settings, particularly in new construction (limits on landscaping, plumbing requirements, and so forth).

Israel faces similar choices, although options are more limited because of its geography. Desalination plants can be constructed, but they yield water at very high costs. Moreover, desalination plants also increase dependence on oil, which is needed to operate the facilities. Wastewater treatment facilities are already used and others are under consideration to recycle some water, but demand management to reduce water use remains the primary tool for meeting projected shortfalls.[18]

Water rights are now a major issue in the Israeli-Arab dialogue. A separate panel is working exclusively on the issue of water and important issues remain unresolved. Until those issues are resolved, options involving transfers from outside Israel also are unlikely until Israel's international situation changes.

Third, complementing the other two factors, is the fact that the urban populations in the two regions have become better educated about water policy trade-offs. Urban users have been forced to examine the system in light of the high costs that resulted from rationing during the recent droughts. As a result of that examination, there is widespread understanding that most water is applied to agriculture, with much of it going to field crops that are relatively low-value crops.

In California, early water politics led to an alliance between agricultural interests in southern California against interests in northern California. Most battles for changing water allocations were between the north and the south. Recently, though, the alliances have changed. Increasingly, the political conflict has shifted to agricultural versus urban uses, with cities pointing to the rising relative value created by water in industrial uses compared to that in agriculture. Similarly in Israel, the political power of the agricultural interests has found increasing opposition among other groups in the matter of water policy.

These factors have made it increasingly difficult to meet growing demands for water through the addition of capacity. Moreover, the increased political power of the environmental groups in California already has forced reallocations of water to the environment that will reduce available supplies.

V. Reform Efforts

Because these problems cannot be answered with traditional solutions, interest has emerged in determining new allocative mechanisms to improve the use of the water that currently is available. Public awareness of the inefficiencies of the current system have bred a large number of groups to reform water policy in the state.[19] Most proposals seek to achieve consensus between agricultural, urban, industrial, and environmental interests. Typically, these consensus approaches call for a mixture of new facilities, conservation ("Best Management Practices"), waste water recycling, increased allocations for environmental protection, and some transfers of water.

16. Total shipments through the Delta need not necessarily be affected by new practices, although the timing of shipments probably will be. Two alternatives that have been considered are (1) a peripheral canal to divert water around the Delta, and (2) storage facilities south of the Delta to bank increased shipments in the wetter winter months to use in the summer months when demands on the system are greater. Both projects face strong political opposition, particularly in the northern part of the state where residents fear that the new facilities would make it possible for increased transfers from the north to the south at the expense of consumers and the environment in the north.

17. Congressional bill H.R. 429, signed into law in October 1992, sets aside 800,000 acre feet for environmental purposes from the CVP, while a proposed ruling by the State Water Resources Control Board (to comply with EPA rulings) may take an additional 300,000 acre feet from other sources.

18. There have been reports that Israel and Jordan currently are considering a new canal from the Red Sea to funnel water to the Dead Sea and generate electricity, but this is viewed largely as a political gesture promoting peace efforts, rather than as a major effort to increase water supplies.

19. In California, these groups include The Three-Way Process, Californians for Water, Committee for Water Policy Consensus, Southern California Water Committee, Farm Water Coalition, and the Bay Delta Oversight Committee, among others.

In most cases, "transfers" are treated as only part of the solution, but water marketing has been rising rapidly to the front of the list of alternatives. Some environmental groups and business groups (such as the Environmental Defense Fund, the Bay Area Economic Forum, the Bay Area Council, and the California Business Roundtable), as well as some of the more prominent urban water districts (most noticeably MWD), pushed hard to bring market forces into water allocation, arguing that the resale of water offers the potential for greater efficiency, with the prospect that nearly all agents can be made better off (Schmidt and Cannon 1991, Mitchell 1993).[20]

These interest groups, were instrumental in obtaining passage of recent federal legislation (H.R. 429), frequently known as the "Bradley-Miller bill," which has strongly embraced the market point of view. The bill, signed into law on October 30, 1992 by President Bush, allows individual contractors to sell up to 20 percent of their allocations without approval by water districts, along with other provisions that allocate water to environmental purposes, create a fund for environmental restoration, and shorten contract periods.

In Israel some steps toward water reform were begun in 1991. These efforts, however, were halted after the change in government following the 1992 election. At present, no significant reforms are under consideration.

This lack of reform efforts in Israel reflects the different social and political interests in the two regions. In California, water policy is increasingly addressed as an economic issue. While arguments still are voiced about the importance of maintaining agriculture in the state, increasingly the discussion has migrated toward economic issues. Arguments opposed to trading emphasize economic dislocations and third party effects, rather than simply relying on statements about the importance of maintaining a way of life for those in the agricultural communities.

In Israel, on the other hand, water remains a strategic resource and the state is vitally concerned with its allocation. As discussed by Wolf and Ross (1992), water policy has been an important consideration in Israel's dealings with its neighbors. For example, according to their analysis, water complicates resolution of the West Bank dispute. The West Bank sits above the Mountain aquifer, and pumping in that region affects supplies to much of central Israel. Under current policies, the Israeli government must approve all pumping from the West Bank. Clearly, should that area no longer fall under Israeli jurisdiction, such control would be jeopardized. Diversions from the Litani River in southern Lebanon also involve strategic interests.

While such strategic considerations do not preclude permission to trade water, they tend to increase the government's interest in monitoring the uses of the resource. Moreover, the kibbutz system has a strong place in the cultural and political structure of Israel. Changes in water policies that might lead to a shift away from agriculture to industrial uses could pose a threat to that system.

Cost of reforms. The speed with which reforms are adopted depends critically on the transitional costs that arise in implementing new policies. Experimentation with additional transfers under the Bradley-Miller legislation should provide strong evidence of the potential gains and disruptions that can result from limited resale of water. By allowing resale of water by CVP contractors, the bill converts water rights into marketable assets, much like electricity from federal projects. Thus, as with electricity, the new structure of rights should encourage marginal transfers among water districts, which may be sufficient to eliminate the need for major new water storage facilities. A key question facing potential reform options is the magnitude of disruptions that such reforms might generate. Would market forces lead to large shifts in water use and to large changes in prices?

Research on California's water system suggests that the quantity of water transferred would be relatively small and the effect on prices to agriculture relatively minor. Howitt, Watson, and Adams (1980) found estimated elasticities of demand for agriculture that were well above those of urban users. Agricultural demand elasticities for water prices in the range of $62 to $87 per acre foot (in 1992 dollars) ranged from -0.98 to -1.5, and prices below this level had even larger elasticities. In contrast, urban users were estimated to have price elasticities close to -0.4 (Vaux and Howitt 1984). Given that agriculture currently consumes somewhere in the range of 80 to 85 percent of the water in California, relatively small percentage reductions in agricultural use resulting from small increases in average water prices would relatively quickly satisfy urban demand: Even a doubling of urban water consumption would reduce agricultural water by only around a fifth from current levels.

In a simulation model embedding these statistics, Vaux and Howitt (1984) estimated that price effects on agriculture and the magnitude of water transferred in California would be relatively small. Using updated figures from Vaux and Howitt's article, Schmidt and Cannon (1991) found that *average* agricultural prices might increase as little as $2.60 per acre foot—from $54.61 to $57.23. Less

20. The extended drought led to the establishment of an emergency water bank in California in 1991. While not a pure market, the bank did provide a mechanism to facilitate transfers from agricultural to urban users, demonstrating the potential for mutually advantageous trade. The water bank, however, is viewed as an emergency measure, and is not generally perceived as a model for marketing water permanently.

than one MAF moved from agriculture to other uses in the simulations. Obviously, some farmers receiving water at well below that price would face a larger increase, but even in those cases, that suggests that those farmers may have the potential to profit from selling more water. These elasticities also are short-run elasticities. Over the longer run, elasticities are likely to be significantly larger as farmers install new technologies that save water.

More recent evidence from Howitt (1991) provides further arguments supporting the low-price impact of a market. According to this research, rice farmers in California could make the same income from selling water at a price of $58 per acre foot (including avoiding production costs), while the break-even price for alfalfa was $114 per acre foot. Given that these commodities, along with irrigated pasture, account for about a third of California's total water use, those prices put a ceiling on the likely level to which prices would rise, since demand by urban areas would be expected to be satisfied well before all of that water would be purchased. Moreover, Howitt found that relatively little water was transferred from agricultural producers of high-value and permanent crops.

Similarly, in Israel, a study by Sadan and Ben-Zvi (1987) examined the implications of allowing water to be traded. They found significant changes occurring in water use across regions, with less used in the northeastern end of the system, and more used in the south. Nevertheless, their study concluded (p.8):

> The findings presented demonstrate the low economic cost of the institutional alternative relative to that provided through new resource development. In the case of Israel, the cost of a given quantity of irrigation water reallocated through institutional change appears to be only half as expensive as that same quantity provided through the implementation of projects for sewage water treatment and recycling, flood control, etc.

While allowing trading would result in some reallocation of resources, and hence some "third party effects" on agricultural communities, concern about such effects must be placed in perspective. Some changes in production and consumption practices would occur, but the indirect effects of those actions on others are likely to be small relative to others that occur regularly in agriculture. For example, the introduction of mechanical tomato harvesters sharply reduced the demand for labor, thus generating third party effects well in excess of those likely to be generated by introduction of a water market (Mitchell 1993).

VI. CONCLUSIONS

Israel and California share similarities not only in their water delivery systems and their institutions, but in their public attitudes. In both cases, water infrastructures have relied heavily on public investments, where costs have been spread widely. Moreover, in both regions political involvement has expanded beyond the construction of facilities to include close controls on allocation and use of the resource. Agriculture has been the biggest beneficiary of past institutional arrangements, typically receiving the bulk of the water and paying lower unit prices for that water.

In both states, serious reform of water policy has proven very difficult. Among the reasons for this difficulty is the ingrained public attitude that because water is "important" it should be allocated administratively. The public seems to believe that this is "more fair" even though actual allocations seem to belie this fairness concept. Despite water's importance in semi-arid areas, policy has opted for this "fairness" over efficiency.

We have argued, however, that concerns over fairness need not preclude trading. As demonstrated in the electric utility industry, it is possible to achieve social policies through differential pricing and through government development of new facilities. Yet, efficiency can be boosted in that system by allowing trading to take place. Fairness can be handled by choosing how to allocate the rights to the resource; efficiency is achieved by granting those rights holders the right to sell to others.

Examination of Israel and California suggests that the willingness to experiment with water reforms—specifically to allow trading—may be increasing in California, while little momentum is apparent in Israel. In part, this may be the result of the trade-off that exists in control of water markets. Direct allocation of water gives tremendous control over development to governmental agencies. The cost of such control, however, is to increase drastically the efficiency losses and encourage poor resource allocation.

In the case of California, momentum is building for increased decentralization of control. This momentum results, in part, from the declining *relative* economic importance of the primary user—agriculture—and the growing importance of environmental values. Since the latter have had the effect of reducing available supplies and making new supplies more difficult to acquire, the efficiency costs implicit in administrative control over water use have risen to the point that other industries and consumers have been forced to address the issue. Moreover, at least in California, evidence suggests that the cost and

disruptions resulting from water trading are not likely to be that large.

In Israel, while economic costs of administrative water allocation also are high, strategic concerns and the political strength of the agricultural sector continue to make reform options politically unpalatable. However, as demand for the water continues to rise with Israel's population, and as other industries become increasingly important relative to agriculture, it is possible that reforms will become more attractive there as well.

REFERENCES

Buchanan, James M., and Gordon Tullock. 1962. *The Calculus of Consent.* Ann Arbor: University of Michigan Press.

Coase, Ronald H. 1960. "The Problem of Social Cost." *Journal of Law and Economics* pp. 1-44.

Department of Water Resources. 1987. *California Water: Looking to the Future.* State of California Department of Water Resources, Bulletin 160-87 (November).

Galnoor, Itzhak. 1978. "Water Policymaking in Israel." *Policy Analysis* 4 pp. 339-367.

Howitt, Richard E. 1991. *Water in California: A Resilient System Under Pressure.* UC/AIC Position Paper No. 91-1. University of California, Davis (April).

———, William D. Watson, and Richard M. Adams. 1980. "A Reevaluation of Price Elasticities for Irrigation Water." *Water Resources Research* 16 (August) pp. 623-628.

Mitchell, David L. 1993. *Water Marketing in California: Resolving Third-Party Issues.* Foster Economics (Bay Area Economic Forum and the Metropolitan Water District of Southern California) (February).

Olson, Mancur. 1982. *The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Rigidities.* New Haven: Yale University Press.

Plaut, Steven E. 1992. *The Suppression of Competition in Israeli Agriculture.* IASPS Policy Studies, No. 11 (April).

Reisner, Marc, and Sarah Bates. 1990. *Overtapped Oasis: Reform or Revolution for Western Water.* Washington, D.C.: Island Press.

Sadan, Ezra, and Ruth Ben-Zvi. 1987. "The Value of Institutional Change in Israel's Water Economy." *Water Resources Research* 23 pp. 1-8.

Schmidt, Ronald H. 1987. "Deregulating Electric Utilities: Issues and Implications." Federal Reserve Bank of Dallas *Economic Review* (September) pp. 13-26.

———, and Frederick Cannon. 1991. *Using Water Better: A Market-Based Approach to California's Water Crisis.* Bay Area Economic Forum (October).

Tahal, 1972. *Israel's Water Economy: Present Status and Outline of Further Development.* Discussion Paper (Tel Aviv, June 1972).

Vaux, Henry J., Jr., and Richard E. Howitt. 1984. "Managing Water Scarcity: An Evaluation of Interregional Transfers." *Water Resources Research* 20 pp. 785-792.

Western Governor's Association Water Efficiency Working Group. 1987. *Water Efficiency: Opportunities for Action* (July 6).

Wolf, Aaron, and John Ross. 1992. "The Impact of Scarce Water Resources on the Arab-Israeli Conflict." *Natural Resources Journal* 32 (Fall) pp. 919-958.