Review

November 1980
Vol. 62, No. 9

The "Rationality" of Survey-Based Inflation Forecasts

R. W. HAFER and DAVID H. RESLER

THE notion that economic agents rationally form their expectations about future economic events has emerged as a critically important hypothesis with profound implications for macroeconomic policy. For example, modern hypotheses relating to the Phillips curve emphasize that it is the departure of *actual* inflation from *expected* inflation that cause any short-run trade-off that may exist between inflation and unemployment. Consequently, empirical tests of many macrotheoretic models require the identification not only of directly observable phenomena, such as inflation and unemployment, but also of expectations or anticipations of these phenomena.

The measurement of generally nonobservable phenomena, such as inflation expectations, poses a difficult challenge in constructing empirical tests for macro models that include such variables. It is first necessary to identify an inflation expectations proxy that is consistent with the assumptions of the underlying model. As a result, tests of theories, such as the natural rate hypothesis, that employ proxy measures for inflation expectations (such as autoregressive procedures) are joint tests of both the underlying theory and the validity of the expectations proxy.

Presumably, autoregressive procedures are used because they are less costly than opinion surveys. When survey-based data on inflation expectations

are readily available, this cost argument loses some of its force. Nevertheless, it is important to determine which of the two measures is appropriate for testing various economic theories; that is, whichever measure conforms most closely to the requirements of the underlying theory becomes the measure of choice. For instance, tests of rational expectations models should first establish that the measures of expectations conform to the criteria of rationality. This paper examines whether one particular set of survey data — the Livingston data — meets specified criteria of rationality.[1]

---

[1]For examples of studies dealing with the measurement and effects of inflation expectations, see John A. Carlson, "A Study of Price Forecasts," *Annals of Economics and Social Measurement* (June 1977), pp. 27-56; Stephen Figlewski and Paul Wachtel, "The Formation of Inflationary Expectations," *Review of Economics and Statistics* (forthcoming); Rodney L. Jacobs and Robert A. Jones, "Price Expectations in the United States: 1947-1975," *American Economic Review* (June 1980), pp. 269-77; Edward Kane and Burton G. Malkiel, "Autoregressive and Nonautoregressive Elements in Cross-Section Forecasts of Inflation," *Econometrica* (January 1976), pp. 1-16; Donald J. Mullineaux, "On Testing for Rationality: Another Look at the Livingston Price Expectations Data," *Journal of Political Economy* (April 1978), pp. 329-36; Douglas K. Pearce, "Comparing Survey and Rational Measures of Expected Inflation: Forecast Performance and Interest Rate Effects," *Journal of Money, Credit and Banking* (November 1979), pp. 447-56; James E. Pesando, "A Note on the Rationality of the Livingston Price Expectations Data," *Journal of Political Economy* (August 1975), pp. 849-58; and Stephen J. Turnovsky, "Empirical Evidence on the Formation of Price Expectations," *Journal of the American Statistical Association* (December 1975), pp. 1441-54.

Tests of Rational Expectations

The hypothesis of rational inflation expectations, pioneered by John Muth, holds that expectations about future inflation are formed in a manner that fully reflects all currently available and relevant information.[2] Stated somewhat differently, the observed rate of inflation differs from the expected rate of inflation only by some random error. Thus, the rationality hypothesis can be stated algebraically as:

$$(1) \quad \pi_t = {}_{t-1}\pi_t^e + u_t,$$

where π_t is the actual rate of inflation during period t, ${}_{t-1}\pi_t^e$ is the rate of inflation expected at time t-1 for period t, and u_t is a random variable with mean zero and variance σ_u^2.[3]

Expressed in this form, i.e., inflation expectations are unbiased estimates of observed inflation, the rationality hypothesis can be tested empirically by estimating the equation,

$$(2) \quad \pi_t = B_0 + B_1 \, {}_{t-1}\pi_t^e + u_t,$$

where ${}_{t-1}\pi_t^e$ represents the survey-based expected inflation rate for period t made at period t-1. The notion of rational expectations, then, corresponds to the joint hypothesis that $B_0 = 0$ and $B_1 = 1$. In addition, u_t should exhibit no evidence of autocorrelation.

Pesando and Figlewski and Wachtel subjected the Livingston expectations series to this test of rationality.[4] Pesando was unable to reject the joint hypothesis using consensus inflation forecasts from each survey for the periods 1959-1969 and 1962-1969. Figlewski and Wachtel, however, were able to reject the null hypothesis using a pooled time series/cross-section sample of 1,864 individual forecasts for the period 1947-1975.

An additional criterion for rationality requires that inflation forecasts be *efficient;* in other words, the process by which inflation expectations are formed should be identical to the process that actually generates observed inflation. Consequently, any evidence suggesting that some of the relevant information set is not being fully (i.e., efficiently) utilized would indicate rejection of rationality. Pesando tested this notion of rationality by hypothesizing that both the expectations of inflation and inflation itself are described by the history of inflation. Mathematically,

this interpretation of rationality can be expressed as:

$$(3) \quad (a) \quad \pi_t = \sum_{i=1}^{n} B_i \, \pi_{t-i} + \mu_{1t}$$

$$(b) \quad {}_{t-1}\pi_t^e = \sum_{i=1}^{n} B_i' \, \pi_{t-i} + \mu_{2t}.$$

Efficiency requires that $B_i = B_i'$ for all i, . . . , n. Pesando, Carlson, and Mullineaux directly tested the efficiency of the Livingston inflation forecasts by estimating equation (3) and then applying an F-test to the sum of the squared residuals.[5] Pesando was not able to reject the efficiency criterion at standard confidence levels for the period 1959-1969. Carlson, using the same time period but a revised version of the Livingston data, found that the inflation forecasts do not satisfy the efficiency criterion.[6]

Mullineaux, on the other hand, demonstrated that the error variances of equations (3a) and (3b) estimated by Pesando and Carlson are not homogeneous. Consequently, the F-test used by Pesando and Carlson is inappropriate.[7] Mullineaux proposed an alternative efficiency test that involves estimating the equation,

$$(4) \quad FE_t = (\pi_t - {}_{t-1}\pi_t^e) = b_0 + \sum_{i=1}^{n} b_i \, \pi_{t-i} + \varepsilon_t$$

where $\varepsilon_t = \mu_{1t} - \mu_{2t}$. The forecast error ($FE_t$) is regressed on past inflation rates known at the time the forecast was made.[8] Efficiency requires that FE_t be

---

[2]John F. Muth, "Rational Expectations and the Theory of Price Movements," *Econometrica* (July 1961), pp. 315-35.

[3]Alternatively, equation (1) can be rewritten as $(\pi_t - {}_{t-1}\pi_t^e) = u_t$; that is, any departure of actual from expected inflation is a random variable with mean zero and variance, σ_u^2.

[4]Pesando, "A Note on the Rationality . . ."; and Figlewski and Wachtel, "The Formation of Inflationary Expectations."

[5]See Pesando, "A Note on the Rationality . . . ." This approach to testing for rationality is generally referred to as a "weak-form" test because it employs *only* information contained in the history of inflation. It should be noted, however, that failure to meet the weak-form requirements of rationality suggests that the forecast would also fail stronger forms of the test. For a discussion of weak-form and other types of tests, see John Rutledge, *A Monetarist Model of Inflationary Expectations* (Lexington: Lexington Books, 1974). In addition, equation (3) does not specify either the exact length of the lag on past inflation or the length of the period over which the inflation is observed. Pesando, Carlson, and Mullineaux each used a 5-period lag on observed 6-month inflation rates. This lag length will also be used in this paper.

[6]Carlson has noted that the numbers published by Livingston have been judgmentally revised. To circumvent this possible source of error, Carlson constructs a forecast series that is based on the actual responses received by Livingston. See Carlson, "A Study of Price Forecasts," for a more detailed discussion of his construction procedures.

[7]The Chow test used by Pesando and Carlson requires that the error terms μ_{1t} and μ_{2t} be independently and identically distributed. If the error terms are *not* identically distributed (homogeneous variances), the Chow test is inappropriate. Mullineaux tests for variance homogeneity by using Bartlett's test statistic and finds that the hypothesis of homogeneous variances is rejected at the five percent level of significance. See Mullineaux, "On Testing for Rationality . . . ," pp. 331-32.

[8]Equation (4) is derived by subtracting equation (3b) from (3a). That is, $b_i = B_i - B_i'$ for all i. Following Mullineaux, equation (4) is estimated with a constant term (b_0) instead of subsuming it into the error structure as Pesando and Carlson did.

unrelated to any information known at the time (t-1) the forecast was formed. In other words, all the informational content of past inflation rates is fully utilized in forming expectations. Thus, the null hypothesis is that $b_0 = 0$ and $b_i = 0$ for all i, . . . , n. In addition, efficiency requires that the error term be serially uncorrelated, or Cov $(\varepsilon_t, \varepsilon_i) = 0$ for $t \neq i$.[9] Using Carlson's version of the Livingston data, Mullineaux was unable to reject the efficiency hypothesis for the period 1959-1969.[10]

Pearce, using Carlson's data set and another test of efficiency, concluded that "the survey respondents did not efficiently use the information in the past history of the Consumer Price Index (CPI) when forming their expectations of inflation."[11] Thus, it appears that efficiency tests of the Livingston inflation expectations data are sensitive to the type of tests used, to the version of the Livingston data used, and to the time period examined.

This article demonstrates that these test results are also sensitive to assumptions about the *length* of the forecast horizon. Therefore, it is particularly important to determine *the actual period over which Livingston respondents are making their forecasts*. The nature of this problem can be illustrated by a careful review of the survey method.

The Forecast Horizon and the Forecast Error

Livingston conducts his survey each spring and fall, requesting respondents to indicate their predictions about a number of economic indicators including the CPI. For example, in the spring survey they are asked to predict what the *level* of the CPI will be in the following December and June. Because the questionnaires are mailed in April and usually are returned in May, two interpretations can be made about the forecast horizon. If, as Carlson assumes, the survey respondents know only the April CPI, then they are implicitly predicting an 8-month rate-of-change (April to December) and a 14-month rate-of-change (April

to June of the following year). Alternatively, Jacobs and Jones argue that a more reasonable assumption is that the respondents actually know or have an accurate estimate of the May CPI.[12] This, of course, means that the forecast CPI implies a 7-month (or 13-month) rate of inflation.

The choice of the forecast horizon can affect the results of the bias and efficiency tests, especially if the forecast is interpreted loosely as a prediction of a steady inflation. Mullineaux and Resler each made this assumption; i.e., they assume that the prediction is a constant rate-of-change for any period within a given forecast horizon.[13] This assumption is often convenient and may not be inappropriate when the investigation focuses on the process that generates the forecast. It may pose problems, however, when efficiency tests, such as those represented by equation (4), are conducted.

Because the survey respondents are, in fact, forecasting an inflation rate over a 7- or 8-month horizon, it is desirable to evaluate equation (4) by calculating the forecast error over that time horizon. For example, FE_t should be calculated by taking the difference between the actual rate of inflation occurring between April (or May) and *December* and the rate of inflation *predicted* for that period. This forecast error should be regressed against lagged inflation rates known to the forecaster as of April (or May). This approach differs from Mullineaux's procedure in which FE_t was computed as of the time the next forecast was made (i.e., October). This approach seems inappropriate for evaluating the efficiency of the forecasts, especially since the forecasts exhibit expectations of accelerating inflation. The next section reevaluates the tests for bias and efficiency in light of these new timing assumptions.

Empirical Results

To investigate the importance that assumptions about the forecast horizon have on tests for bias and

---

[9]It should be noted that, although the heterogeneous variance problem that plagued the Chow tests of Pesando and Carlson is alleviated here, the procedure employed does require the maintained hypothesis of independent errors.

[10]Mullineaux also found that for the data set used by Pesando (i.e., inflation forecasts inferred from the originally published versions of Livingston data), the hypothesis of efficiency is rejected.

[11]Pearce, "Comparing Survey and Rational Measures . . . ," p. 451. Pearce statistically analyzes the forecast errors obtained by using either the Livingston forecasts or forecasts generated from a continuously updated moving average model [MA(1)] of the monthly CPI series.

[12]Jacobs and Jones, "Price Expectations in the United States: 1947-1975."

[13]This essentially requires that inflation forecasts are linear. Thus, changes from one point to another within the forecast horizon will not be distinguishable. If, however, inflation expectations are not linear over different time horizons (e.g., 6 or 8 months), then the assumption of a steady rate of inflation prediction is vitiated. The fact that the 14-month forecasts are greater than the 8-month forecasts in 38 out of 40 observations from 1959-1978 suggests that the assumption of a constant rate of inflation within the 8- or 14-month periods may not be appropriate. See Mullineaux, "On Testing for Rationality," fn. 3. See also, David H. Resler, "The Formation of Inflation Expectations," this *Review* (April 1980), pp. 2-12.

efficiency (and hence rationality), the three alternative forecast horizons discussed in the preceding section are utilized in direct empirical comparisons. Based on these forecast horizons, three forecast error series are calculated and employed in the efficiency tests reported below. To reiterate, these alternative FE_t series are determined by assuming an April-October forecast horizon (Mullineaux), a May-December forecast horizon (Jacobs-Jones), and an April-December forecast horizon. All tests use Carlson's version of the Livingston data (i.e., sample average CPI forecasts from which the expected inflation rate is generated). To facilitate a comparison with previous research, the following sample periods are used: 1959-1969, 1959-1978, and 1959-1978 excluding the 1971-1973 period of price controls of various phases.[14]

To test for bias in the inflation forecasts, equation (2) is estimated and an F-test on the joint hypothesis that $B_0 = 0$ and $B_1 = 1$ is conducted for each of the alternative forecast horizons.[15] The F-values calculated for this test are presented in table 1, and allow rejection of the null hypothesis at the 1 percent level, irrespective of the sample period chosen. This result contrasts directly with Pesando's but is consistent with the findings of Figlewski and Wachtel, who found the Livingston data to be biased.[16] An examination of the individual coefficients, B_0 and B_1, indicated that the joint hypothesis is rejected primarily because B_1 exceeds unity for all the sample periods. Nevertheless, the results indicate a tendency for B_1 to decline toward unity as more recent observations are added to the sample, suggesting that forecasters gradually adjusted to the accelerating inflation of the 1960s and early 1970s.[17]

Table 2 presents additional information on the accuracy of the inflation expectations series. Although the root-mean-squared error and mean error statistics

---

Table 1

Bias Test for "Short-Run" Inflation Forecasts[1] F-Values

Forecast horizon	Period		
	1959-1969	1959-1978	1959-1978[2]
April-October	15.242	15.723	14.401
May-December	12.660	15.487	14.411
April-December	28.367	18.144	17.439
	F(2,20)	F(2,38)	F(2,35)
Critical 5%	3.49	3.25	3.27
F-values 1%	5.85	5.21	5.27

[1]Test based on joint hypothesis that $B_0 = 0$ and $B_1 = 1$ in equation (2').

[2]This period excludes the 1971-1973 price control years.

---

vary only slightly between forecast horizons, the Theil statistics indicate that the fraction of forecast error due to bias is reduced somewhat by using the May-December horizon. It is interesting to note that, of all of the horizons examined, the April-December assumption continually yields statistics suggesting greater problems with bias than variance or covariance in the forecasts.[18]

Although unbiased forecasts satisfy one criterion for rationality, it is common to find properties of bias in other non-survey-based inflation forecasts. For instance, Lombra and Moran note that, while the Federal Reserve Board staff's forecasts of nominal GNP are unbiased, its forecasts of GNP's real and inflation components show evidence of systematic errors.[19]

It is possible that inflation forecasts can show evidence of systematic bias yet still be characterized as

---

[14]This truncated 1959-1978 sample period was chosen to exclude observations of forecasts errors that occurred during the period of wage and price controls. It seems reasonable that forecasters would have encountered considerably more difficulty in forecasting inflation during this period, since the controls were applied unevenly and gradually relaxed at unpredicted intervals.

[15]To facilitate computation of the appropriate F-statistics, equation (2) was modified slightly. Specifically, subtracting $_{t-1}\pi_t^e$ from each side of (2) produces:

$$(2')\quad \pi_t - {}_{t-1}\pi_t^e = B_0 + (B_1 - 1)_{t-1}\pi_t^e + u_t.$$

The null hypothesis then implies that the estimated slope and intercept of equation (2') be jointly equal to zero.

[16]Pesando, "A Note on the Rationality . . . " and Figlewski and Wachtel, "The Formation of Inflationary Expectations."

[17]In studies of the *process* by which inflation forecasts are generated, more definitive evidence indicates that this process has changed over time. For more detail about this evidence, see Donald J. Mullineaux, "Inflation Expectations and Money

---

Growth in the United States," *American Economic Review* (March 1980), pp. 149-161, and Resler, "The Formation of Inflation Expectations."

[18]For a description of this methodology, see Henri Theil, *Applied Economic Forecasting* (Amsterdam: North Holland Publishing Co., 1971), pp. 26-32.

[19]Raymond Lombra and Michael Moran, "Policy Advice and Policy Making at the Federal Reserve," *Carnegie-Rochester Conference Series on Public Policy 13*, 1980, p. 20. For evidence that other forecasts similarly underestimate inflation and over-estimate real output, see V. Zarnowitz, "An Analysis of Annual and Multiperiod Quarterly Forecasts of Aggregate Income, Output, and the Price Level," *Journal of Business* (1979), p. 133.

Table 2
Analysis of the "Short-Run" Forecast Errors[1]

Forecast horizon assumption	Sample period	RMSE	Mean error	Theil statistics		
				U^m	U^s	U^e
April-October	1959-69	1.383	0.911	0.434	0.360	0.206
	1959-78	2.151	1.324	0.379	0.226	0.394
	1959-78[2]	2.053	1.270	0.383	0.223	0.394
May-December	1959-69	1.344	0.858	0.408	0.347	0.246
	1959-78	2.317	1.414	0.372	0.252	0.375
	1959-78[2]	2.214	1.356	0.375	0.252	0.373
April-December	1959-69	1.307	0.934	0.513	0.344	0.143
	1959-78	2.101	1.355	0.416	0.210	0.374
	1959-78[2]	1.962	1.261	0.413	0.203	0.384

[1]RMSE is the root-mean-squared error, U^m is the Theil bias coefficient, U^s the variance coefficient, and U^e the covariance coefficient.

[2]Omits the 1971-1973 price control years.

"weakly" rational in the sense that the forecasters efficiently utilize all information contained in the history of inflation. To implement this *efficiency* test, FE_t is calculated for each forecast horizon and used to estimate equation (4).

Because acceptance of the efficiency hypothesis in the present context requires that $b_i = 0$ for all i (i=1, ..., n) and that the estimated relationships indicate no evidence of serial correlation, the statistics of primary interest are the reported F-values and the Durbin-Watson and Durbin-h statistics. The reported F-value is pertinent for testing the joint hypothesis that all the b_i (i = 1, ... ,5) are concurrently zero. Both the Durbin-Watson and Durbin-h statistics test for the presence of serial correlation. Although the Durbin-Watson statistic is usually appropriate, Durbin has shown that the h statistic is more efficient when the set of independent variables includes a lagged dependent variable.[20] Because Mullineaux has interpreted equation (4) as containing a lagged dependent variable, both statistics are reported.

Ordinary least squares estimates of equation (4), using the alternative FE_t series and sample periods, are presented in table 3. These results differ consider-

ably from those of Mullineaux, and they highlight the importance of specifying the time period over which FE_t is calculated. If FE_t is evaluated at the end of the period over which the respondents were forecasting inflation (e.g., December), the efficiency hypothesis is rejected in all but one instance. The results for the three different time periods are now discussed in greater detail.

Turning first to the 1959-1969 period, the reported F-statistic for the May-December and the April-October forecast horizons indicates that the efficiency criterion is satisfied. Recalling that the April-October horizon corresponds to the assumption made by Mullineaux, these results are essentially consistent with his. The Durbin-h statistic for the April-October horizon, however, indicates the presence of negative serial correlation, even though the Durbin-Watson statistic falls within the indeterminate range.[21] Since

[20]See James Durbin, "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors are Lagged Dependent Variables," *Econometrica* (May 1970), pp. 410-21.

[21]For purposes of comparison, Mullineaux's estimation results are presented here:

$$(\pi_t - {}_{t-1}\pi_t^e) = -0.232 + 0.237\pi_{t-1} - 0.051\pi_{t-2} + 0.251\pi_{t-3}$$
$$(1.91) \quad (1.44) \quad (0.27) \quad (1.36)$$

$$+ 0.050\pi_{t-4} + 0.083\pi_{t-5}$$
$$(0.25) \quad (0.48)$$

$$\bar{R}^2 = 0.102, h = 1.89, F = 1.48.$$

The difference between Mullineaux's results and those in table 3 may well be due to the use of different computer algorithms. As such, the difference between the Durbin-h values may not be representative of true differences in the respective residual processes.

Table 3
Efficiency Test Results[1]

Forecast horizon	Coefficients[2]						Summary statistics[3]				
	b_0	b_1	b_2	b_3	b_4	b_5	$\bar{R}^2$	D.W./h	S.E.E.	F	F*(.05, .01)
				1959 - 1969							
April-October	−0.244 (0.46)	0.244 (1.48)	−0.049 (0.26)	0.254 (1.37)	0.050 (0.26)	0.083 (0.48)	0.11	2.61/−2.26	1.00	1.54	
May-December	−0.493 (1.02)	0.193 (1.19)	0.114 (0.61)	0.215 (1.10)	0.114 (0.58)	0.041 (0.22)	0.24	2.25/−0.91	0.92	2.36	2.85, 4.44
April-December	−0.345 (0.88)	0.218 (1.79)	0.051 (0.36)	0.295 (2.16)	0.019 (0.13)	0.061 (0.47)	0.38	1.85/0.43	0.74	3.52	
				1959 - 1978							
April-October	0.695 (1.53)	0.397 (2.82)	0.003 (0.02)	−0.100 (0.55)	−0.261 (1.44)	0.102 (0.71)	0.20	2.14/−0.97	1.54	2.92	
May-December	0.442 (1.01)	0.435 (3.29)	0.133 (0.81)	−0.113 (0.64)	−0.436 (2.43)	0.200 (1.40)	0.37	1.98/0.12	1.47	5.68	2.49, 3.51
April-December	0.717 (1.74)	0.368 (2.88)	0.035 (0.22)	−0.058 (0.35)	−0.362 (2.21)	0.160 (1.23)	0.26	1.77/1.01	1.40	3.76	
			1959 - 1978 (Omitting 1971-73)								
April-October	0.649 (1.44)	0.300 (2.06)	0.062 (0.35)	−0.052 (0.28)	−0.257 (1.41)	0.082 (0.56)	0.14	2.21/−1.22	1.51	2.16	
May-December	0.414 (0.96)	0.340 (2.43)	0.175 (1.06)	−0.051 (0.29)	−0.412 (2.31)	0.155 (1.07)	0.34	2.01/−0.05	1.50	4.43	2.54, 3.73
April-December	0.668 (1.69)	0.269 (2.11)	0.098 (0.63)	−0.009 (0.06)	−0.361 (2.26)	0.140 (1.10)	0.24	1.80/0.90	1.33	3.14	

[1]Test results based on equation (4).

[2]Values in parentheses represent absolute values of t-statistics.

[3]$\bar{R}^2$ is the coefficient of determination corrected for degrees of freedom; D.W. is the Durbin-Watson statistic; h is the Durbin-h statistic; S.E.E. is the standard error of the equation; F is the calculated F-value to test the joint hypothesis that all b_i (i = 1, . . . , 5) equal zero; and F* represents the relevant critical F-value.

efficiency requires no serial correlation among the residuals, the hypothesis of efficiency for the April-October horizon remains unresolved. Unlike these two forecast horizons, however, the results based on using the April-December assumption clearly permit rejection of the efficiency hypothesis.[22]

In contrast to the results for the 1959-1969 period, the hypothesis of efficiency is unambiguously rejected at the 5 percent level for each forecast horizon examined during the entire 1959-1978 sample period. The hypothesis is also rejected at the 1 percent level for the May-December and April-December horizon

periods. Based on these test results, the period from 1959-1978 does not appear to be one in which Livingston forecasters, on average, efficiently utilized the information contained in the history of observed inflation rates.

Similarly, when the period of wage price controls is excluded, the efficiency criterion is not satisfied if the forecast error is calculated at the end of the forecast period (e.g., in December). For instance, when the forecast error is measured at the end of the period over which the forecast is made, the F-test permits a rejection of the efficiency hypothesis at the 5 percent level.[23] The efficiency hypothesis is not re-

[22]It should be recalled that the April-December forecast horizon does not require the special assumptions necessary to construct the competing forecast error series. We *know* that Livingston supplies the April CPI to the survey recipients and specifically asks for their *December* CPI forecast.

[23]The efficiency hypothesis cannot be rejected, however, at the 1 percent level when the 8-month (April-December) forecast horizon is employed.

jected *only* when the forecast error is evaluated in October (as in Mullineaux).

Efficiency of the 12-Month Forecasts

Most previous analyses of the Livingston inflation forecasts focus on the short-run (8-month) forecasts. Because the respondents are asked at each survey date to predict the level of the CPI for the following December and June, the forecasts embody both an 8-month and a 14-month (long-run) prediction of the inflation rate. This section examines the rationality of the 14-month forecasts.

The methodology used here slightly modifies the approach used for the 8-month forecasts. Specifically, the lagged inflation rates in equation (4) are now interpreted as occurring over 12-month periods (again, observed in either April or October). This assumption requires that the estimation of these equations for the 14-month forecasts be modified.

Because the forecasts are made at 6-month intervals, this new interpretation means that the first lagged term in equation (4) contains information that overlaps from the previous period, if all available observations are included in the estimation procedure. Such overlapping observations may introduce serial correlation into the equation.[24] To avoid this problem, separate estimations of equations (2) and (4) are made for each semiannual observation of the 14-month forecast; that is, each sample period is split into two data sets, one consisting only of the June forecasts and the other consisting only of the December forecasts. With these modifications, equations (2) and (4) are estimated for the three time periods used in the previous section.

The analysis first examines the 14-month forecasts for bias. F-statistics were computed from the regressions of equation (2) for each semiannual forecast series over each sample period. These F-values, reported in table 4, again indicate that the forecasts are biased. Table 5 provides the statistics for Theil's analysis of the forecast errors. These results also show that 33-54 percent of the forecast error is due to bias. Nevertheless, as with the "short-run" forecasts, the portion due to bias declines as new data are added.

The efficiency test is then applied to the 14-month forecast errors. The forecast errors are consistently

Table 4

Bias Test for 14-Month Inflation Forecasts[1] F-Values

Sample period	June forecast	December forecast	Critical F (.05, .01)
1959-1969	16.130	20.800	4.26, 8.02
1959-1978	9.533	5.188	3.55, 6.01
1959-1978[2]	10.599	4.592	3.63, 6.23

[1]Test based on joint hypothesis that $B_0 = 0$ and $B_1 = 1$ in equation (2').

[2]Omits the 1971-1973 price control years.

measured as of the end of the period over which the forecast was made. The F-statistics and the Durbin-Watson statistics for these equations are reported in table 6.[25] In contrast to the 8-month (April-December) inflation forecasts, the results for the 14-month forecasts *do not* permit rejection of the efficiency hypothesis. Because halving the sample period severely reduces the degrees of freedom, these results should be interpreted with considerable caution. Nevertheless, the F-statistics suggest that the errors in the 14-month forecasts are not correlated with observations of past inflation available at the time the forecast was made. The Durbin-Watson statistics, however, indicate that the hypothesis of no serial correlation can neither be rejected nor accepted. Thus it appears that, based on the F-test, the 14-month forecasts comply with the efficiency criterion.

These contrasting results for the 8-month and 14-month forecast horizons cast some doubt on the findings that the Livingston forecasts are not formed efficiently. This disparity may indicate that forecasters are better able to anticipate longer-term movements in economic variables, such as inflation, relative to explaining the short-term vagaries of the time series. For instance, if the actual rate of inflation is accelerating within the 14-month period, the forecaster may be able to forecast efficiently the overall rate of change but not be able to forecast the rate within shorter sub-periods.

---

[24]Introduction of serial correlation tends to bias the efficiency test toward rejecting the null hypothesis. Recall that an additional criterion for efficiency is that the estimation be free of autocorrelation.

[25]The Durbin-h statistic is not appropriate for small samples (n<30). On this point, see J. Johnston, *Econometric Methods,* 2nd ed. (New York: McGraw-Hill, 1971).

Table 5
Analysis of 14-Month Forecast Errors[1]

Forecast horizon assumption	Sample period	RMSE	Mean error	Theil statistics		
				U^m	U^s	U^c
June	1959-69	1.120	0.824	0.540	0.337	0.123
	1959-78	1.964	1.298	0.436	0.208	0.356
	1959-78[2]	2.022	1.383	0.468	0.222	0.310
December	1959-69	1.182	0.782	0.438	0.474	0.088
	1959-78	2.085	1.190	0.326	0.198	0.477
	1959-78[2]	1.976	1.133	0.329	0.194	0.477

[1]RMSE is the root-mean-squared error; U^m is the Theil bias coefficient; U^s the variance coefficient; and U^c the covariance coefficient.

[2]Omits the 1971-1973 price control years.

Table 6
Efficiency Test Results: 14-Month Forecasts[1]

Sample period	N	June forecasts		December forecasts		F*(.05, .01)
		F	D.W.	F	D.W.	
1959-69	11	0.426	1.38	1.344	2.25	5.05, 10.97
1959-78	20	1.049	1.88	2.029	1.87	2.96, 4.69
1959-78[2]	18	0.875	1.34	1.993	2.30	3.11, 5.06

[1]N is the respective sample size; F is the calculated F-statistic; D.W. is the Durbin-Watson test statistic; and F* represents the relevant critical F-value.

[2]Omits the 1971-1973 price control years.

Summary

This paper has reexamined the rationality of the inflation forecasts contained in the Livingston survey data by emphasizing that the inflation forecast error should be calculated in a manner consistent with the forecast horizon used by the survey respondents. Specifically, empirical tests for bias and efficiency of the forecasts were employed to determine the effect that changes in the assumption about the forecast horizon have on the conclusions of previous investigations. The test for bias indicated that, regardless of the forecast horizon or the sample period used, the Livingston forecasts exhibited characteristics of bias.

The "efficiency" test suggested by Mullineaux was also employed. These test results indicate that over the period, 1959-1969, only one forecast horizon (April-December) could be judged unambiguously inefficient. When the 1959-1978 period is examined, however, the results for each forecast horizon allow rejection of the efficiency hypothesis. When the period of wage-price controls is deleted from this sample period, only the April-October forecast horizon is judged efficient.

These findings imply that conclusions regarding the forecast efficiency (and, therefore, rationality) of the Livingston inflation expectations are sensitive to the period over which the forecast error is evaluated. Because the survey respondents are asked specifically to predict the level of the CPI for the following June or December, it seems appropriate that tests of

efficiency be formulated to measure the forecast error *only* after the actual value of the predicted CPI becomes known. When this approach is used in conjunction with the assumption of either a May-December or April-December forecast horizon, the results indicate that the forecasters did not efficiently use the information available at the time of the survey in five out of six samples. This conclusion contrasts sharply with that reached when the forecast error is calculated at the time the forecasts are made (i.e., April or October).

Finally, evidence about the bias and efficiency of the 14-month forecasts indicates that these longer forecasts are efficient, even though, like the 8-month forecasts, they are apparently biased. Although the apparent disparity in the efficiency tests between the "short-" and "long-run" forecasts is somewhat puzzling, it suggests that the forecasters are more efficient at predicting longer term inflation trends than short-term movements in the series.

The evidence presented here indicates that Carlson's sample average forecasts of the rate of CPI inflation in the Livingston data do not conform to two criteria of rationality. Consequently, the use of these data in empirical investigations of rational expectations models appears to have serious limitations. In addition, the observation that these survey-based inflation expectations fail to conform to rationality criteria suggests that adjustments in expectations evolve slowly. This further implies that, even if inflation forecasts are ultimately rational, fully anticipated short-run monetary policy actions may have important economic effects since inflation expectations adapt slowly. These and other possible implications of the apparent non-rationality of survey-based expectations deserve further study.

Monetary Aggregates as Monetary Indicators

KEITH M. CARLSON and SCOTT E. HEIN

THE monetary aggregates are being relied upon more and more as indicators of the thrust of monetary policy actions on aggregate economic activity.[1] To be useful as a monetary indicator, a monetary aggregate should satisfy at least two criteria. First, it must be sensitive to policy actions taken by the Federal Reserve — such as open market operations and changes in reserve requirements, the discount rate, and Regulation Q ceilings; it must not be sensitive to influences other than Federal Reserve actions. If the monetary aggregate is responsive to nonpolicy forces, it will provide erroneous signals as to the thrust of monetary policy.[2]

Second, a monetary aggregate should be both consistently and predictably related to the pace of economic activity. If it is not, changes in the monetary aggregate will not "indicate" what will happen to aggregate economic activity as a result of actions currently being taken by monetary authorities.

Early this year, the Federal Reserve Board announced a redefinition of the monetary aggregates. In some cases, the differences between the old and new money measures are quite substantial. While the relationship between the old monetary aggregates and economic activity has received much attention in the economic literature, the usefulness of the new monetary aggregates as monetary indicators has yet to be examined in detail. This article reports some results bearing on this issue.

The analysis focuses primarily on the relationship of the new M1A, M1B, and M2 measures to economic activity. To provide historical continuity, the results are compared with those derived from analyses of the old M1, M2, and M3 aggregates.

THE NEW MONETARY AGGREGATES

Components of the new M1A, M1B, and M2 monetary aggregates are listed in table 1.[3] M1A is identical

---

[1]For a general discussion of monetary indicators, see Albert E. Burger, "The Implementation Problem of Monetary Policy," this *Review* (March 1971), pp. 20-30.

[2]This criterion explains why many argue against the use of market interest rates as monetary indicators. See Albert E. Burger, "The Implementation Problem . . . ," where he argues that market interest rates are poor monetary indicators because they are sensitive to nonpolicy impulses, such as factors that affect the demand for credit.

[3]For a detailed description of the new monetary aggregates, see R. W. Hafer, "The New Monetary Aggregates," this *Review* (February 1980), pp. 25-32; or Thomas D. Simpson, "The Redefined Monetary Aggregates," *Federal Reserve Bulletin* (February 1980), pp. 97-114.

to old M1, except that it excludes demand deposits due to foreign commercial banks and official institutions. The new M1B aggregate, a broader transaction measure, is equal to M1A, except that it includes newly developed interest-bearing *transaction* deposits. These latter deposits include negotiable order of withdrawal (NOW) accounts, automatic transfer system deposit (ATS) accounts, and credit union share drafts. NOW accounts were legalized in certain New England states early in the 1970s, and such legalization will extend nationwide as of December 31, 1980.[4] Commercial banks have been permitted to offer individual ATS accounts since November 1, 1978.

Chart 1 presents compounded annual rates of change of old M1, M1A, and M1B for the period II/1959 through IV/1979.[5] The chart shows that the exclusion of demand deposits held by foreign commercial banks and institutions has had little effect on the growth rates of the monetary aggregates. Growth rates of new M1A closely resemble those of old M1. Furthermore, the growth rates of M1A and M1B differ little prior to early 1974 and, although M1B growth usually exceeds that of M1A over the period I/1974 through III/1978, the disparity between these aggregates is quite small. It is only after the nationwide introduction of ATS accounts in late 1978 that the growth rates of these new aggregates show any marked divergence.

While the new M1A and M1B measures are similar in scope to old M1, the new M2 measure is quite different from old M2. In fact, the new M2 measure is more closely related to the old M3, which included savings and small time deposits of thrift institutions; old M2 did not include such deposits. Because the monetary aggregates are no longer differentiated on the basis of institutional considerations, old M2 does not have a counterpart among the new measures.

As shown in table 1, there is essentially only one component of the old M3 measure — large time deposits (other than large negotiable CDs) at commercial banks and thrift institutions — that is not included

Table 1
The New Monetary Aggregates

Component	M1A	M1B	M2
Currency in circulation	X	X	X
Demand deposits at commercial banks and thrift institutions, exclusive of deposits due to foreign commercial banks and official institutions	X	X	X
NOW and ATS accounts and credit union share drafts		X	X
Overnight RPs			X
Savings deposits at commercial banks and thrift institutions			X
Small time deposits (less than $100,000) at commercial banks and thrift institutions			X
Overnight Eurodollar deposits issued by Caribbean branches of member banks and held by U.S. nonbank residents			X
Money market mutual fund shares			X

in the new M2 measure. On the other hand, a number of the changes that have been made make new M2 even more comprehensive than old M3. In addition to the interest-bearing transaction deposits included in M1B, the new M2 measure also includes overnight RPs at commercial banks, money market mutual funds, and overnight Eurodollar deposits issued by Caribbean branches of member banks and held by U.S. nonbank residents.[6]

Chart 2 depicts the compounded annual rates of change of new M2, old M2, and old M3. Growth rates of the new M2 and old M3 aggregates were similar from the II/1959 through II/1973 period; growth rates of old M2, on the other hand, generally were much slower than these aggregates. The similarity in the growth rates of old M3 and new M2 breaks down in late 1973, however, when overnight RPs, money market mutual funds, and the overnight Eurodollar deposit component of new M2 became increasingly popular.

[4]For a description of the New England experience with NOW accounts, as well as a discussion of how their legalization will affect other parts of the country, see William N. Cox III, "NOW Accounts: Applying the Northeast's Experience to the Southeast," *Economic Review* of the Federal Reserve Bank of Atlanta (September/October 1980), pp. 4-10; and Patrick J. Lawler, "NOW Accounts in the Southwest: A Break for Consumers, an Entry from S&L's, and a Test for Banks," *Voice of the Federal Reserve Bank of Dallas* (October 1980), pp. 1-8.

[5]The historical series for the new monetary aggregates begins in I/1959.

[6]Timothy Q. Cook and Jeremy G. Duffield, "Short-Term Investment Pools," *Economic Review* of the Federal Reserve Bank of Richmond (September/October 1980), pp. 3-23. The authors have recently argued that there are many other investment pools, similar to money market mutual funds, which should be included in the new M2 measure.

Chart 1
Compounded Annual Rates of Change of
M1(old), M1A and M1B

Percent

- M1(old)
- M1A
- M1B

1959 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 1980

Latest data plotted: M1(old)- 4th quarter 1979; Others-2nd quarter 1980
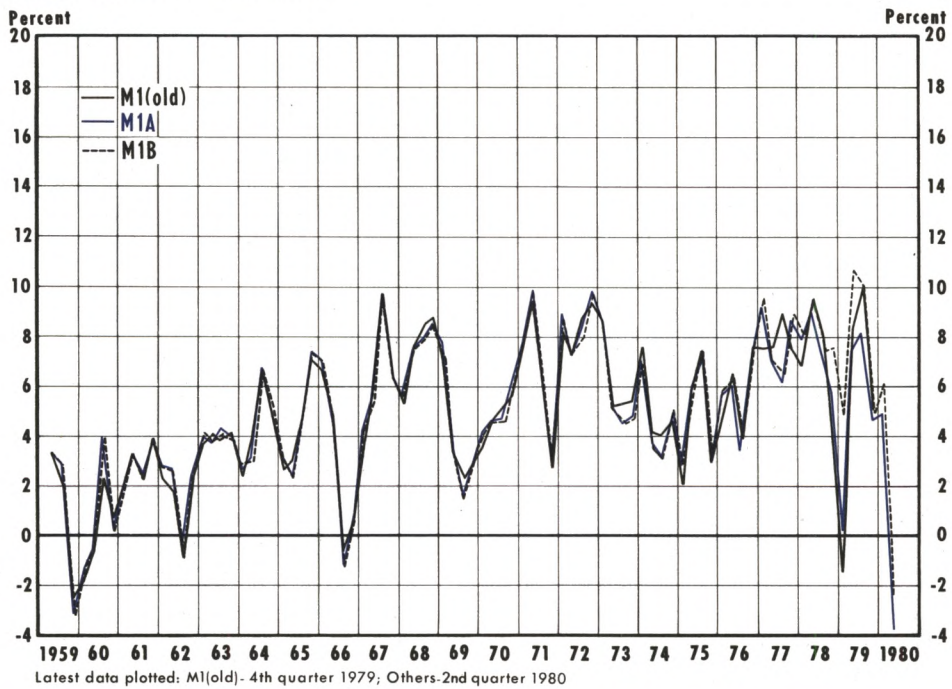
Chart 2
Compounded Annual Rates of Change of
M2(new), M2(old) and M3(old)

Percent

- M2(new)
- M2(old)
- M3(old)

1959 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 1980

Latest data plotted: M2(new)-2nd quarter 1980; Others-4th quarter 1979
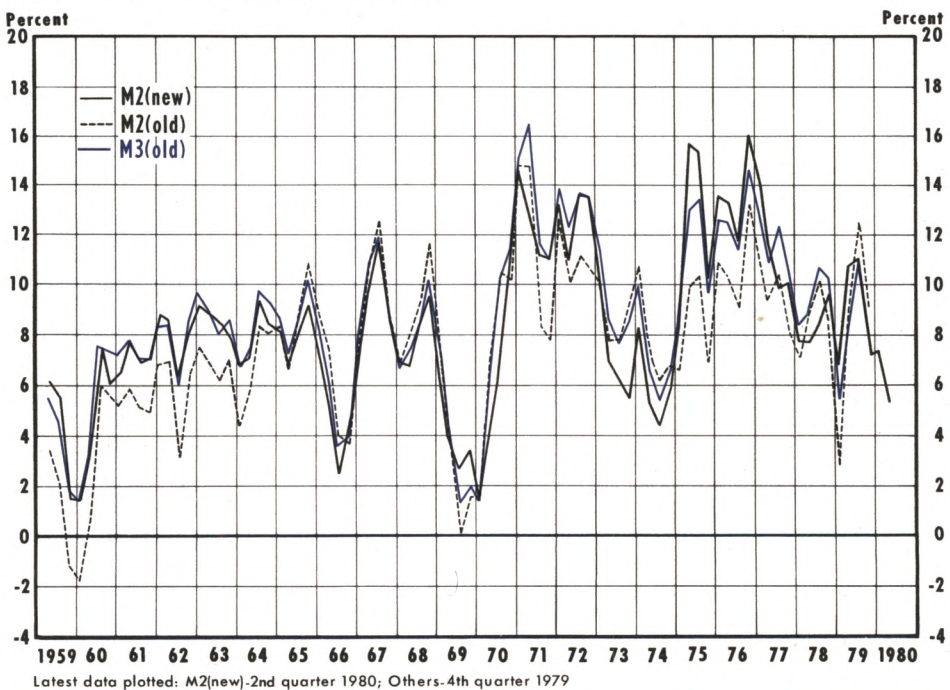
Chart 3
Compounded Annual Rates of Change of M2(new) and M1B

Percent

Shaded areas represent periods during which the three-month treasury bill rate was at least 100 basis points above the ceiling rate on commercial bank savings deposits.
Latest data plotted: 2nd quarter 1980
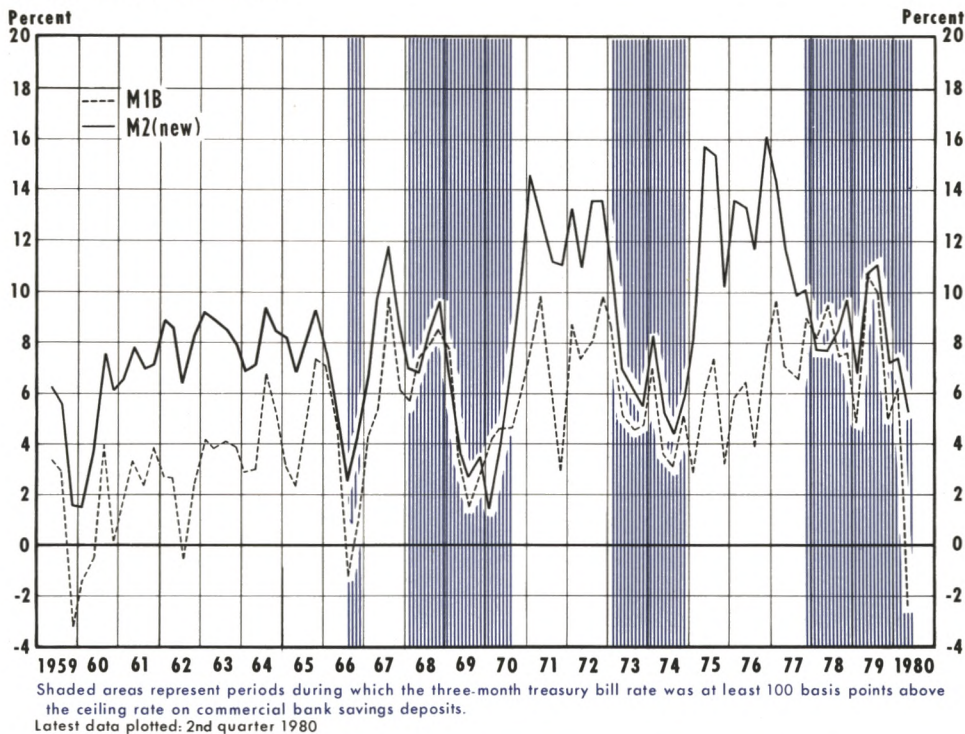
Finally, chart 3 presents the compounded annual rates of change of the new M1B and M2 aggregates. This chart illustrates the differential growth rates of narrow versus broad money measures.[7] Note the difference in average growth rates; new M2 growth is usually above that of M1B. The average growth rate of new M2 over the II/1959 through IV/1979 period is 8.4 percent, compared to 5.0 percent for M1B.

The differential between the two growth rates sometimes varies. The chart indicates a definite pattern in the relative growth rates. Over the periods II/1959-IV/1965, III/1970-I/1973, and I/1975-I/1978, growth rates of new M2 are substantially above those of M1B. In the intervening periods, the differential between growth rates of these two aggregates is very small.

Historical experience indicates that the growth rate of the broad money stock measure is sensitive to the differential between market interest rates and Regulation Q ceilings. This is clearly indicated by the

shaded areas in chart 3, which depict periods of two quarters or more during which the three-month treasury bill rate was at least 100 basis points above the ceiling rate on commercial bank savings deposits.[8] Redefining this broader monetary aggregate has not made it insensitive to nonpolicy influences. Nonpolicy factors that affect the supply or demand for credit and, as a result, change market interest rates will clearly influence the growth of new M2 just as they affected the growth of old M2 and M3. The sensitivity of new M2 to such nonpolicy factors thus reduces its usefulness as an indicator of monetary policy actions.

THE RELATIONSHIP BETWEEN ECONOMIC ACTIVITY AND THE MONETARY AGGREGATES

The relationship between economic activity and the new monetary aggregates is investigated with

---

[7]M1A is excluded for simplification purposes; prior to late 1978, quarterly growth rates of M1A were very similar to those of M1B (see chart 1). Further, while only the new aggregates are shown, old M1 and M2 display a similar pattern.

[8]The chart indicates that the most recent period of disintermediation, IV/1977-II/1980, has not had the same effect in reducing new M1B as observed in previous periods of disintermediation. However, at least part of this phenomenon is explained by the rapid growth of overnight RPs and Eurodollar deposit holdings and, more recently, by money market mutual funds.

reference to nominal GNP. Nominal GNP is chosen because this is the apparent channel by which monetary policy variables directly affect the economy.[9] The general form of the relationship to be estimated is:

$$(1) \qquad \dot{Y}_t = C + \sum_{i=0}^{f} m_i \dot{M}_{t-i} + \sum_{i=0}^{g} e_i \dot{E}_{t-i} + \mu_t$$

where $\dot{Y}$ is the compounded annual growth rate of nominal GNP, $\dot{M}$ is the compounded annual growth rate of the given monetary aggregate, $\dot{E}$ is the compounded annual growth rate of high-employment expenditures, and μ is a random error term.[10] This relationship is estimated using the new M1A, M1B, and M2 aggregates and the old M1, M2, and M3 measures. The relationships are estimated with the ordinary least squares estimation technique.

The investigation subjects the six different relationships to a number of statistical tests. The strategy is first to find the optimal lag structure for the different relationships over the sample period, III/1962 through III/1977. After investigating the in-sample stability of the relationships and the likelihood of simultaneous equation bias problems, these estimated relationships are then used to project nominal GNP over the post-sample period, IV/1977 through IV/1979, to determine which relationship would have yielded the most accurate forecasts for this period. This period was chosen because of the divergent growth rates for the various aggregates, as shown in the preceding charts.

Sample Period Relationships

The first concern in estimating the general relationship given in equation (1) is to determine the appropriate values of f and g, the number of lags on the monetary and fiscal variables. Lag values of 0, 4, and 8 were considered for each of the six relationships. Interestingly enough, F-tests for each of the equations indicated that the appropriate lag value was

4 for each of the separate monetary aggregates, as well as for the fiscal variable.

Table 2 provides the sample period coefficient estimates and summary statistics for the six different equations, where the relationships are estimated with ordinary least squares and four lags on the fiscal and monetary variables are assumed. There is very little difference between the sample period fit provided by the various aggregates. In all cases, the standard error of the estimating equation (SEE) is less than one-third the size of average GNP growth over the sample period (9.61 percent).

While the pattern of the distributed lag effects of both the fiscal and monetary variables is similar across equations, the size of the coefficients is clearly dependent on the comprehensiveness of the monetary aggregate employed. In general, the more comprehensive the aggregate, the smaller the size of any lagged monetary coefficient. The sum of the money coefficients is close to 1.0 for both M1A and M1B.[11] On the other hand, the sum of the money coefficients for new M2 is close to 0.7. Regardless of the aggregate used, the sum of the high-employment expenditures coefficients is close to zero.

Stability Tests

A question to be considered with these estimation results is whether the relationships reported in table 2 are structurally stable (i.e. whether the regression coefficients change significantly with time). The hypothesis of structural stability was investigated with the use of the Chow test. The formal hypothesis tested is whether the regression coefficients estimated for the III/1967 through IV/1969 sample period differ significantly from those obtained for the same equation in the I/1970 through III/1977 period. The null hypothesis is that the coefficients are equal in each of these periods. The midpoint of the sample was chosen as the breakpoint because it maximizes the power of the test.[12]

Table 3 lists the F-statistics for each of the various equations. None of the cases considered provide evi-

---

[9]See Milton Friedman, "A Theoretical Framework for Monetary Analysis," in *Milton Friedman's Monetary Framework: A Debate with His Critics*, ed. R. J. Gordon (University of Chicago Press, 1974), pp. 1-63; and Charles R. Nelson, "Recursive Structure in U.S. Income, Prices, and Output," *Journal of Political Economy* (December 1979), pp. 1307-27.

[10]This relationship is similar to the original Andersen-Jordan equation. Such a relationship has been estimated more recently by Keith M. Carlson, "Money, Inflation, and Economic Growth: Some Updated Reduced Form Results and Their Implications," this *Review* (April 1980), pp. 13-19. Usually, the relationship is estimated assuming that the lag coefficients lie along a polynomial of a given degree. No such constraints are imposed here.

[11]The results reported for the narrow aggregates are similar to those found by Keith M. Carlson, "Money, Inflation and Economic Growth . . . ," where a third degree polynomial with tail constraints was employed in the estimation.

[12]See John U. Farley, Melvin Hinich, and Timothy W. McGuire "Some Comparisons of Tests for a Shift in the Slopes of a Multivariate Linear Time Series Model," *Journal of Econometrics* (Volume 3, No. 3, 1975), pp. 297-318.

Table 2
Relationships Between GNP and The Monetary Aggregates

$$\dot{Y}_t = c + \sum_{i=0}^{4} m_i \dot{M}_{t-i} + \sum_{i=0}^{4} e_i \dot{E}_{t-i} + \mu_t$$

(Sample Period, III/1962-III/1977; absolute value of t-statistic in parenthesis)

| | Monetary aggregates | | | | | |
| | Old | | | New | | |
	M1	M2	M3	M1A	M1B	M2
C	2.94 (2.00)	0.56 (0.32)	0.90 (0.54)	2.35 (1.56)	2.20 (1.49)	1.30 (0.85)
m_0	0.58 (2.97)	0.38 (2.03)	0.16 (0.81)	0.61 (3.20)	0.60 (3.18)	0.13 (0.74)
m_1	0.02 (0.10)	0.14 (0.59)	0.26 (0.90)	0.03 (0.13)	0.05 (0.24)	0.39 (1.67)
m_2	0.20 (0.83)	0.12 (0.52)	0.08 (0.27)	0.32 (1.36)	0.31 (1.35)	−0.07 (0.28)
m_3	0.56 (2.28)	0.43 (1.82)	0.49 (1.68)	0.36 (1.53)	0.38 (1.62)	0.44 (1.85)
m_4	−0.54 (2.67)	−0.19 (0.99)	−0.30 (1.48)	−0.35 (1.75)	−0.35 (1.79)	−0.22 (1.28)
e_0	0.04 (0.89)	0.05 (0.98)	0.08 (1.74)	0.05 (1.18)	0.05 (1.19)	0.08 (1.95)
e_1	0.12 (2.67)	0.10 (2.15)	0.13 (2.80)	0.12 (2.55)	0.12 (2.59)	0.12 (2.80)
e_2	−0.07 (1.54)	−0.07 (1.54)	−0.07 (1.61)	−0.08 (1.81)	−0.08 (1.82)	−0.08 (1.88)
e_3	0.02 (0.42)	0.01 (0.23)	0.02 (0.34)	0.02 (0.35)	0.01 (0.31)	0.02 (0.43)
e_4	xx[1] (0.02)	−0.02 (0.43)	−0.01 (0.28)	0.01 (0.13)	xx[1] (0.04)	−0.01 (0.23)
R^2	0.49	0.46	0.49	0.48	0.49	0.51
SEE	2.96	3.05	2.95	2.98	2.98	2.90
DW	1.78	1.87	2.00	1.82	1.89	2.03

[1]less than 0.005

Table 3

Calculated F-Statistics for Test of Break In Relationships
(III/1962-IV/1969 vs. I/1970-III/1977)

| | Monetary aggregates | | | | | |
| | Old | | | New | | |
	M1	M2	M3	M1A	M1B	M2
$F_{(11,39)}$[1]	1.52	1.11	0.64	1.64	1.62	0.61

[1]The 5 percent critical level is 2.05; the 10 percent critical level is 1.73.

dence to reject the null hypothesis at traditional levels of significance.[13]

Simultaneous Equation Bias Tests

A further question with regard to the estimation results reported in table 2 is whether or not they are subject to significant simultaneous equation bias. Equations such as those reported in table 2 can be estimated reliably with ordinary least squares methods only if the independent variables are exogenous. A major criticism of equations of this type is that the monetary aggregates are not exogenous with respect to GNP.[14]

Sims has recently suggested a test to examine whether the independent variables in a distributed lag relationship, such as equation (1), can be said to be statistically exogenous.[15] The test procedure involves adding *leading* values of the independent variables to the basic distributed lag equation. If the regression coefficients of the *leading* values of the

---

[13]A break in the relationship in I/1974 was also considered. With the exception of the new M2 relationship, there is evidence, at traditional levels of significance, to suggest a break in all the relationships. With regard to the inability to reject the stability of the new M2 relationship, it should be noted that *none* of the separate subperiod money coefficients differed from zero.

The fact that all other equations break is evidence of the specification error. There appear to be two likely candidates for omitted variables. First, none of the relationships include a variable to capture the impact of the oil shock which occurred near 1974. Second, there is no variable to capture a shift in money demand if, as many argue, money demand shifted in 1974. (For example, see Stephen M. Goldfeld, "The Case of the Missing Money," *Brookings Papers on Economic Activity* (3:1976), pp. 683-730.)

Since we are primarily concerned with the coefficients on the money variables, either of these specification errors will cause a problem only to the extent that the excluded variable is correlated with the independent variables. It is only when such correlation exists that the estimated coefficients will be biased. Regardless of whether either or both of the above specification errors exist, it is unlikely that this bias problem will result. Both of the suggested specification errors resulted because shock variables were excluded. For evidence of the "shock" view of money demand, see R. W. Hafer and Scott E. Hein, "The Dynamics and Estimation of Short-Run Money Demand," this *Review* (March 1980), pp. 26-35. By definition, these shock variables should not be correlated with the included independent variables. The out-of-sample simulation results to be reported later in this paper indicate that there is little evidence of a significant bias in these simulations.

[14]See Frank de Leeuw and John Kalchbrenner, "Monetary and Fiscal Actions: A Test of Their Relative Importance in Economic Stabilization — Comment," this *Review* (April 1969), pp. 6-11.

[15]Christopher A. Sims "Exogeneity and Causal Ordering in Macroeconomic Models," in *New Methods in Business Cycle Research: Proceedings from a Conference* (Federal Reserve Bank of Minneapolis, 1977), pp. 23-44.

---

independent variable are not different from zero, the null hypothesis of exogeneity is supported. On the other hand, statistical significance of *leading* coefficients suggests that simultaneous equation bias problems would result if the equation were estimated with ordinary least squares.

To test for the presence of simultaneous equation bias, four leads on both the fiscal and monetary variables were added to the basic equation as follows:

$$(2) \qquad \dot{Y}_t = C + \sum_{i=0}^{4} m_i \dot{M}_{t-i} + \sum_{i=0}^{4} e_i \dot{E}_{t-i} + \sum_{i=1}^{4} m_i' \dot{M}_{t+i}$$

$$+ \sum_{i=1}^{4} e_i' \dot{E}_{t+i} + \mu_t.$$

Since the Sims test depends crucially on the statistical significance of regression coefficients, every effort was made to assure the absence of serially correlated error terms. This was accomplished by following Sims' recommendation of filtering the data prior to estimation. In most cases, the filter employed was the first order linear filter (1-KL), where L is the lag operation and K is a constant. The value of K was determined by iterating over values from 0 to 1, at intervals of 0.1. The first value of K which yielded no evidence of a relationship between the contemporaneous residual and residuals lagged, first two and then four periods, was chosen as the appropriate value.[16]

This search procedure removed the problem of serially correlated disturbances in all relationships except that using old M1. In this case, the fourth lagged residual always remained statistically significant in an autoregressive error structure in the residuals. Thus, in the case of old M1, the filter employed was (1-KL⁴).

Table 4 lists the F-statistics testing the null hypotheses; (1) $m_i' = 0$ for i = 1, 2, 3, 4; (2) $e_i' = 0$ for i = 1, 2, 3, 4; and (3) $m_i' = e_i' = 0$ for i = 1, 2, 3, 4. In none of the cases considered were F-statistics large enough to reject the null hypothesis at the 5 percent level, thus suggesting the absence of any simultaneous equation bias problems associated with the estimation results reported in table 2.[17]

---

[16]A similar search procedure was employed by Yash P. Mehra and David E. Spencer, "The St. Louis Equation and Reverse Causation: The Evidence Reexamined," *Southern Economic Journal* (April 1979), pp. 1104-20.

[17]This conclusion is somewhat different than that obtained by Mehra and Spencer, "The St. Louis Equation. . . ." In estimating a relationship similar to equation (1), they found evidence of simultaneous equation bias problems. However, their study differed in three important ways. First, the only

Table 4
F-Statistics for Simultaneous Equation Bias Tests

| Null hypothesis[1] (degrees of freedom) | Monetary aggregates | | | | | |
| | Old | | | New | | |
	M1	M2	M3	M1A	M1B	M2
$m_i' = 0$ (4,42)[2]	1.62	0.57	0.85	0.23	0.25	2.20
$e_i' = 0$ (4,42)[2]	0.22	0.47	0.12	0.14	0.13	0.27
$m_i' = e_i' = 0$ (8,42)[3]	0.94	0.37	0.44	0.23	0.23	1.18
Value of K	0.10	0.10	0.10	0.20	0.20	0.10

[1]i = 1, 2, 3, 4 in all cases.

[2]The 5 percent and 10 percent critical values for F(4,42) are 2.60 and 2.09, respectively.

[3]The 5 percent and 10 percent critical values for F(8,42) are 2.17 and 1.82, respectively.

Two qualifications to this conclusion are required. These qualifications concern the regressions employing old M1 and new M2. While the F-statistics reported in table 4 do not allow the rejection of the null hypothesis at the 5 percent level, there were *individual* lead money coefficients in these two cases that were different from zero at certain levels of significance; thus, there is some evidence to reject the null hypothesis at lower significance levels. For example, in the case of old M1, the regression coefficient on the one-quarter lead of money was 0.64. The t-statistic associated with this individual coefficient was 2.32, indicating that the estimate was statistically different from zero at the 5 percent level. In this regard, there is some evidence of "reverse causation" — an increase in economic activity "causing" an increase in future money growth.[18] This result generates some concern about the regression estimates reported for the equation using old M1 in table 2.

It is interesting to note that the redefinitions of the monetary aggregates, although not directly concerned with this simultaneity problem, have done much to resolve it. None of the individual leading money coefficients were close to being statistically different from zero when the M1A aggregate was employed. Together, these findings suggest that the simultaneous

equation bias, to the extent it exists, is due to the inclusion of demand deposits held by foreign institutions or commercial banks.

In the case of new M2, the coefficient on the money variable led two quarters was –0.50; and its absolute t-statistic of 1.83 was significantly different from zero at the 10 percent level. In addition, the joint hypothesis that all leading money coefficients are zero had to be rejected at the 10 percent level. This again suggests the possibility of a simultaneous equation bias problem. However, it is important to recognize that the problem does not appear to be a result of a *positive* association between current economic activity and future money growth, as traditionally suggested. Rather, in this case, this regression coefficient suggests that current economic activity is *negatively* associated with new M2 growth two quarters in the future.[19]

This negative relationship should not come as a surprise in light of the evidence of the impact of disintermediation on new M2 growth. An increase in economic activity, by causing market interest rates to rise above Regulation Q ceilings, will be associated, other things being equal, with a reduction in future new M2 growth.

In summary, it appears that the redefinitions of the

---

monetary variable they consider is the monetary base. Second, they include high-employment receipts, as well as high-employment expenditures, in their relationship. Finally, they focus on a different time period (I/1952-IV/1974).

[18]More formally, if one were willing to use the 25 percent significance level, the null hypothesis that the leading M1 coefficients are equal to zero must be rejected.

[19]In this regard, it is to be noted that when old M3 is used, the coefficient on money variable led two quarters is also negative. However, the coefficient is not different from zero even at the 10 percent level. Thus, it appears that including overnight RPs, overnight Eurodollars, and money market mutual funds in new M2 has compounded the simultaneity problem.

Table 5
Summary Measures of Out-of-Sample (IV/1977-IV/1979) GNP Predictions

| | Monetary aggregates | | | | | |
| | Old | | | New | | |
	M1	M2	M3	M1A	M1B	M2
Mean error	−2.17	−2.67	−2.60	−1.61	−0.02	−2.57
Root mean squared error	4.05	4.55	4.62	4.04	3.68	5.03
Fraction of error[1]						
due to bias	0.29	0.34	0.32	0.16	0.00	0.26
due to variation	0.29	0.36	0.31	0.20	0.47	0.17
due to covariation	0.42	0.29	0.38	0.64	0.53	0.56

[1]Need not sum to unity as a result of rounding.

monetary aggregates have removed possible problems associated with simultaneity as far as the narrow transaction aggregates are concerned. However, there still remains a question concerning simultaneity with regard to the more comprehensive measure.

Prediction Results

How well do the relationships presented in table 2 simulate nominal GNP over the IV/1977 through IV/1979 period? Table 5 indicates that the equation using the new M1B aggregate performs the best in simulating GNP growth over this period, regardless of the criteria considered. The strength of this equation is most evident in the lack of bias in the predictions. The other aggregates underpredict GNP growth over this period, on average, by approximately 2.5 percent. In comparison, the average prediction error for M1B is a trivial −0.02 percent.

It is also appropriate to note that the bias in prediction errors is smaller for new M1A than for old M1. Removing demand deposits held by foreign commercial banks and institutions did not reduce the variance of forecast error; it did, however, reduce the average error and the bias in the forecast.

The fact that the more comprehensive monetary aggregates (old M2, old M3, and new M2), which include savings deposits subjected to Regulation Q ceilings, *underpredict* GNP growth by more than the transaction aggregates is again consistent with the

view that disintermediation has adversely affected the growth of these deposits. The whole period from IV/1977 through IV/1979 has been characterized by market interest rates well above Regulation Q ceilings. This has led to a relative slowing in the growth of these regulated deposits. As a result, equations using these aggregates have underpredicted economic activity since IV/1977.

SUMMARY

The monetary aggregates were redefined early this year. The purpose of this article was to examine these new aggregates in terms of their usefulness as monetary policy indicators. Two criteria for judging the usefulness of the monetary aggregates as indicators were suggested. First, to serve as an indicator, the aggregate should reflect the policy actions of the monetary authority and not be highly sensitive to nonpolicy influences. Second, the aggregate should be consistently and predictably related to economic activity.

Although the first criterion was not considered formally, examination of the rates of change of the new monetary aggregates indicated that redefining M2 did not remove the influence of nonpolicy forces. In particular, the movement of market interest rates relative to Regulation Q ceilings has had an adverse effect on new M2 growth (relative to the narrowly defined aggregates), as it did with the old M2 and M3 aggregates.

The second criterion was examined more extensively by regressing nominal GNP growth on the growth of the various monetary aggregates and a fiscal variable (growth rates of high-employment expenditures). These relationships were checked for structural stability, simultaneous equation bias, and out-of-sample prediction accuracy. Of the new monetary aggregates, only M2 showed any evidence of simultaneous equation bias. This problem is felt to be closely related to the impact of Regulation Q ceilings. In out-of-sample simulations, M1B performed better than any of the other new aggregates analyzed, indicating that it had a closer relationship to economic activity than did the other new aggregates.

In light of the criteria suggested for judging the usefulness of the new monetary aggregates as mone- tary indicators, M1B was thus found to best satisfy these requirements. It appears to be relatively insen- sitive to nonpolicy influences (a characteristic it shares with M1A), and it is more predictably and consistently related to movements of nominal GNP than M1A or new M2.

On the other hand, new M2 was found to be par- ticularly unreliable as a monetary indicator. Growth in this aggregate was found to be sensitive to non- policy forces. While proposed actions under the Finan- cial Institutions Deregulation and Monetary Control Act of 1980 should eventually resolve this type of problem, new M2 growth will likely remain a poor monetary indicator in the seven-year transition period, especially in light of the absence of any reliable his- torical relationship with economic activity.

EIGHTH FEDERAL RESERVE DISTRICT

HEAD OFFICE, BRANCH, AND OTHER CITIES OF OVER 10,000 POPULATION

Kirksville

QUINCY

Hannibal

Jacksonville

Moberly

Mexico

COLUMBIA

Fulton

ST. LOUIS

Centralia

Jefferson City

Mount Vernon

Bedford   Seymour

Vincennes

Washington

Madison

Frankfort

LOUISVILLE

Rolla

Murphysboro

EVANSVILLE

Marion

Fort Knox

Danville

Fort Leonard Wood

Carbondale

OWENSBORO

Elizabethtown

SPRINGFIELD

Cape Girardeau

Madisonville

Sikeston   Paducah

CLARKSVILLE-HOPKINSVILLE

Bowling Green

Glasgow

Poplar Bluff

Mayfield

Murray

Union City

FAYETTEVILLE-SPRINGDALE

Paragould

Dyersburg

Humboldt

Jonesboro   Blytheville

Jackson

FORT SMITH

Russellville

Conway

Forrest City

MEMPHIS

Corinth

LITTLE ROCK

West Helena

Stuttgart

Hot Springs

Helena

Oxford

Tupelo

PINE BLUFF

Clarksdale

Camden

Cleveland

TEXARKANA

Greenwood

Columbus

Magnolia

Greenville

Starkville

El Dorado

Legend

- ✪ Head Office of the FRB, St. Louis
- ✴ Branch Offices of the FRB, St. Louis
- ■ Standard Metropolitan Statistical Areas

- ▣ Places over 50,000
- ◼ Places of 40-50,000
- ◉ Places of 30-40,000
- ● Places of 10-20,000
- ○ Places of 20-30,000

Population is based on the 1970 census.

District States

Ind.

Mo.

Ky.

Ark.

Tenn.

Miss.

— State Boundaries
— District Boundary

Scale in Miles   0   32   64   96

Federal Reserve Bank of St. Louis DECEMBER 1975