

An Analysis of RSQE Forecasts: 1971-1992

E. PHILIP HOWREY*

The purpose of this paper is to evaluate the accuracy of ex ante econometric model forecasts of four key macroeconomic variables: real GNP growth, the rate of price inflation measured by the GNP deflator, the civilian unemployment rate, and the Treasury Bill rate. Annual forecasts produced by the Research Seminar in Quantitative Economics (RSQE) based on the Michigan Quarterly Econometric Model of the U.S. Economy are compared with quasi ex ante forecasts from a four-variable vector autoregressive (VAR) model. Statistical tests of the equality of forecast error variances as well as univariate and multivariate forecast encompassing-type tests are conducted. The forecast error variance comparisons indicate that for three of the four variables the RSQE forecasts are more accurate than the VAR forecasts and for one of the variables (real GNP growth) only slightly less accurate. The forecast encompassing-type tests indicate that the RSQE forecasts contain information not contained in the VAR forecasts and, conversely, that VAR forecasts contain information not included in the RSQE forecasts. The scope for improving RSQE forecasts by combining them with VAR forecasts is rather limited, however. (JEL C32, C53, E17)

I. Introduction

The Research Seminar in Quantitative Economics (RSQE) at the University of Michigan was established in 1951. One of the first research projects of the seminar was the construction of an econometric model of the U.S. economy. A version of what has come to be known as the Klein-Goldberger model was used in November 1952 to produce the first RSQE forecast. Each year since then, a macroeconomic forecast has been presented at the Annual Conference on the Economic Outlook which is held in November.

The purpose of this paper is to analyze the recent RSQE forecasting record. Such an analysis is of interest for several reasons. Macroeconometric forecasting models have been under attack for a number of years. As Sims [1980, p. 1] notes,

"though large-scale statistical macroeconomic models exist and are by some criteria successful, a deep vein of skepticism about the value of these models runs through that part of the economics profession not actively engaged in constructing or using them."

Lucas and Sargent [1981, p. 303] are less refrained:

"the track record of the major econometric models is, on any dimension other than very short-term unconditional forecasting, very poor.... macroeconometric models were subjected to a decisive test in the 1970s.... the inflationary bias on average of monetary and fiscal policy in this period, according to all of these models, should have produced the lowest average unemployment rates for any decade since the 1940s. In fact, as we know, they produced the highest unemployment rates since the 1930s. This was econometric failure on a grand scale."

* The University of Michigan, Ann Arbor.

A systematic examination of the forecasting record may shed some light on whether this skepticism and pessimism are warranted.

More recently, interest has centered on the combination of forecasts from alternative models as a way to obtain improved forecasts. Nelson [1984] found that a composite forecast obtained by combining a single equation, autoregressive, integrated, moving average model for real GNP growth with five different macroeconomic models produced more accurate forecasts than the econometric models alone for two-, three-, or four-quarter ahead forecasts for the period 1976-82. Lupoletti and Webb [1986], using forecasts from Data Resources Inc., Chase Econometrics, and Wharton Econometric Forecasting Associates for the period 1970-83, found statistically significant evidence that the combination of econometric model forecasts with VAR forecasts resulted in an improved composite forecast. Both of these studies point to the conclusion that simple univariate and multivariate time series models contain useful information for forecasting that is often overlooked by well-known commercial forecasters, at least over the 1970-83 time period.

Various types of forecast encompassing tests have been proposed in the literature, including both univariate and multivariate procedures. However, most of the empirical results that have been reported are based on univariate analysis. Fair and Shiller [1989, 1990], for example, confine their analysis to GNP forecasts obtained from a set of models. Both univariate and multivariate tests are examined in this paper to see if such tests provide useful information for model evaluation.

The analysis in this paper is consistent with two themes that were emphasized in Professor Harold Hotelling's work on prediction. In his paper with Holbrook Working, Hotelling [1929, p. 73] argued that "the probable error of a trend is always appropriate, and furthermore, it is frequently necessary if sound conclusions are to be formed." An examination of the RSQE forecasting record provides a basis for assigning a measure of probable error to the forecasts. If econometric forecasts were accompanied by such probable error estimates, unrealistic expectations of forecast precision might be avoided.

Hotelling [1940] also took up the issue of selection of one variate from a set of variates that are available to predict a variable of interest. The statistical question dealt with in that paper [p. 278] was "the distribution function suitable for testing the hypothesis that there is no real difference between any pair of the correlations of x_1, \dots, x_p with y ." This theme of model selection and comparison with an emphasis on statistical testing was explicitly advocated by Hotelling [1942, p. 72]: "The discriminant function ought not be used for any such purpose of prediction, however, until it is verified that the assignment into classes that it makes is at least as good as the result of assignment by tossing a coin." The distribution of the test statistic appropriate for a test of this hypothesis had been developed earlier by Hotelling [1931]. These ideas are rather clear precursors of forecast encompassing test procedures.

The organization of the remainder of this paper is as follows. The forecasts which are analyzed and the models on which they are based are briefly described in the next section. The results of root mean squared error comparisons are summarized in the section that follows. Univariate and multivariate encompassing-type tests are then examined. The paper concludes with a summary of the results.

II. Models and Forecasts

The Research Seminar forecasts examined in this paper are the forecasts presented at the Annual Conference on the Economic Outlook held in November. These forecasts are based on the Michigan Quarterly Econometric Model (MQEM) of the U.S. Economy. This model has evolved over time from the initial Klein-Goldberger model [Klein and Goldberger, 1955], which was an annual model consisting of 20 equations. Today, MQEM contains over 200 equations, roughly half of which are stochastic. For a brief overview of the structure of MQEM, see Howrey and Hymans [1995]. The dynamic characteristics of the IS and LM curves implicit in MQEM are described by Green et al. [1991].

The RSQE forecasts released in November of each year are true *ex ante* forecasts in the sense that they are based on quarterly data available through the third quarter of the year. They are not, however, pure model forecasts in the sense that they reflect both model projections and the wisdom and expertise of the forecasting team.¹ Thus, the comparisons that follow are not model evaluation tests per se, but rather an evaluation of a model-based forecasting procedure.

For purposes of comparison, a simple four-variable VAR model has been constructed using quarterly data. The four variables are real GNP growth (Y), the rate of inflation as measured by the GNP deflator (P), the unemployment rate (U), and the 90-day Treasury bill rate (R).² The number of lags included in VAR models is often determined on an ad hoc basis. As shall be seen, this can lead to an overparameterization of the model and a decrease in forecast accuracy. Order selection criteria can be used to guard against overparameterization of the VAR model.³

The values of alternative order selection criteria are shown in Table 1, based on quarterly data for the period 1959:2-1992:4. The values of the four criterion functions shown in Table 1 are Akaike's final prediction error criterion (FPE), Akaike's information criterion (AIC), Schwartz's Bayesian information criterion (BIC), and the Hannan-Quinn criterion (HQ). The estimated order of the VAR is the value of the number of lags that minimizes the criterion function. FPE, AIC, and HQ select a VAR of order 3, while BIC selects a VAR of order 2. It is well known that FPE and AIC tend to overestimate the order of the VAR. Thus, comparisons with the VAR2 forecasts will be highlighted in what follows. However, comparisons with higher-order VAR model forecasts will also be examined to document the decrease in forecast accuracy that can accompany overparameterized models.

In an attempt to ensure comparability of the information sets on which the forecasts are based, the VAR model was estimated using quarterly observations from 1954:2 through the third quarter of the year prior to the forecast year. The same information was used to

¹ For an attempt to isolate the role of judgment in RSQE forecasts, see Donihue [1993].

² Sims [1980] used a six-variable VAR model which included real GNP, the GNP deflator, the unemployment rate, the nominal wage rate, and the money supply. Fair and Shiller [1990] added the 3-month Treasury bill rate to this list of variables. Lupoletti and Webb [1986] used the monetary base, real GNP, GNP deflator, manufacturing capacity utilization index, and the Treasury bill rate in their VAR model.

³ For a discussion and Monte Carlo simulation study of alternative criteria for determining the order of a VAR model, see Lütkepohl [1985].

TABLE 1
VAR Model Selection Criteria

LAG	FPE	AIC	BIC	HQ
0	1.34	6.88	6.88	6.88
1	0.79	-0.30	0.06	-0.15
2	0.52	-0.72	0.00	-0.43
3	0.43	-0.92	0.16	-0.48
4	0.45	-0.87	0.56	-0.29
5	0.50	-0.77	1.02	-0.04
6	0.49	-0.80	1.35	0.07
7	0.57	-0.66	1.84	0.35
8	0.60	-0.63	2.24	0.53

estimate the coefficients of the VAR model as was available at the time the RSQE forecast was prepared. A 5-quarter forecast was generated and the quarterly forecasts were then converted to average annual rates for comparison with the RSQE forecasts. However, the VAR model forecasts are not completely comparable to the RSQE forecasts for several reasons. First, the VAR models were estimated using the currently available data as of the December 1994 National Income and Product Accounts release, not the data that were available at the time the forecasts would have been made. Thus, the VAR database includes revised data not available at the time the RSQE forecasts were released. Second, the VAR models and forecasts use 1987 constant-dollar GNP and the corresponding GNP deflator. The RSQE forecasts are based on the data available at the time the forecasts were prepared, which involve different base years. For example, prior to December 1975, GNP and the GNP deflator were measured in 1958 dollars. The 1976-85 RSQE forecasts were based on 1972 dollar GNP, and the most recent (1986-92) forecasts were based on 1982 dollar GNP. It is extremely difficult, after the fact, to construct true *ex ante* forecasts. Such forecasts are frequently referred to as quasi *ex ante* forecasts.

III. Comparisons of Forecast Accuracy

The primary series analyzed in this paper are plotted in Figures I-IV. Figure I shows the actual annual rate of growth of real GNP (in 1987 dollars) for 1971-92, the RSQE forecast of the growth rate announced at the Outlook Conference in November of the previous year, and the projected growth rate from a vector autoregressive model of order 2 (VAR2).

The similarity of the RSQE and VAR2 forecasts for all four variables is readily apparent. With the possible exception of the Treasury bill rate, it is difficult to distinguish between the RSQE and VAR2 forecasts on the basis of simple visual inspection of the graphs.

FIGURE I
ANNUAL RATE OF GROWTH OF REAL GNP

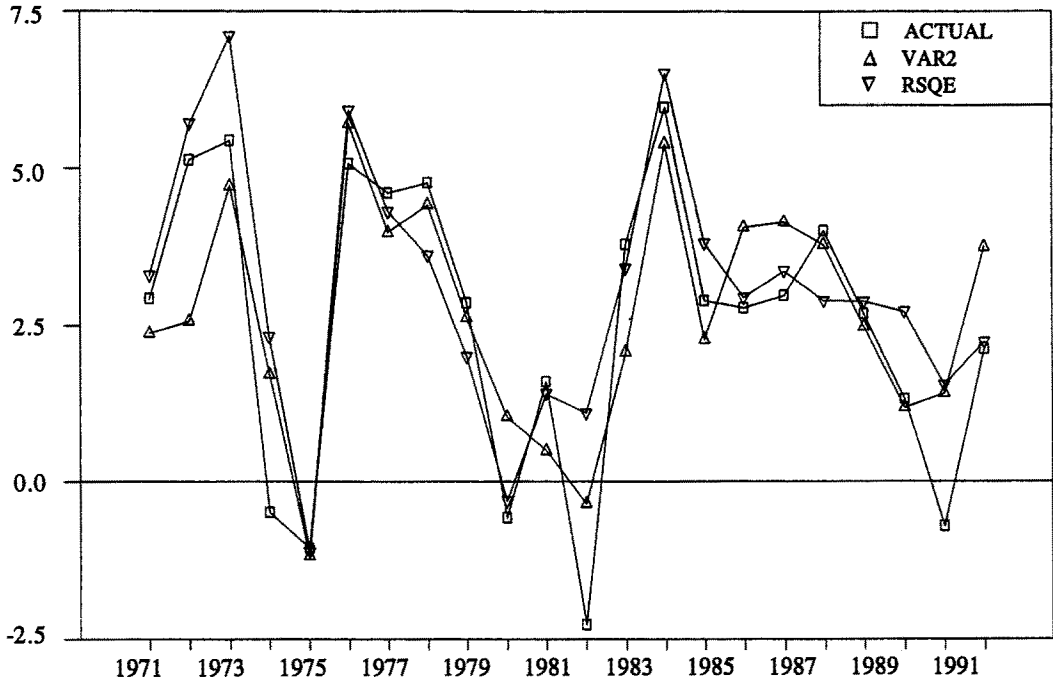


FIGURE II
ANNUAL RATE OF INFLATION - GNP DEFLATOR

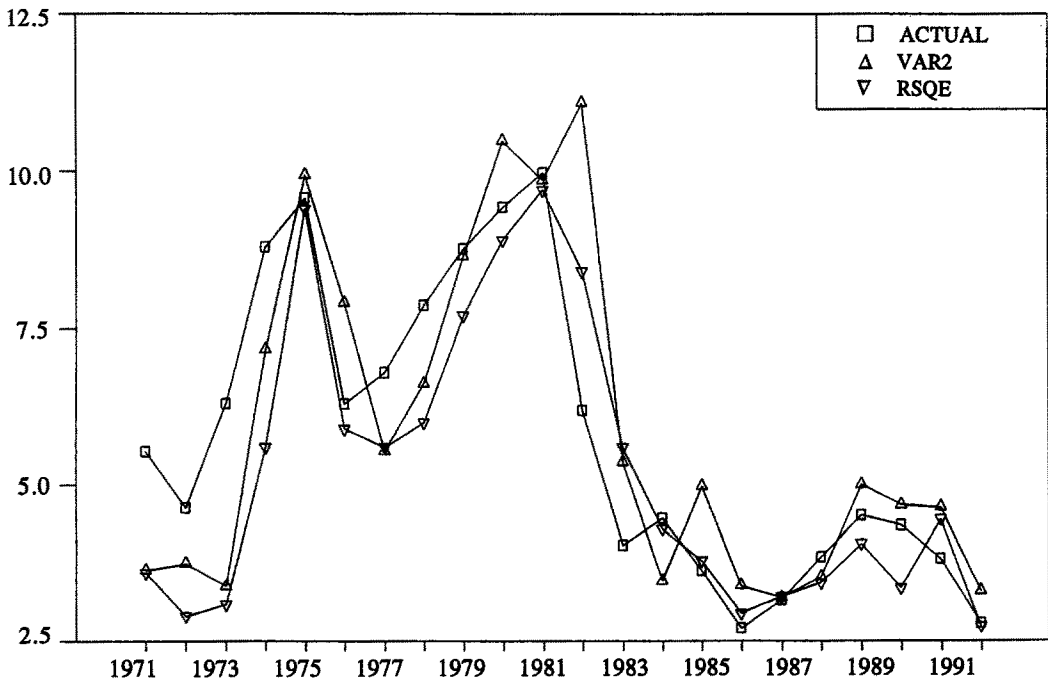


FIGURE III
TREASURY BILL RATE

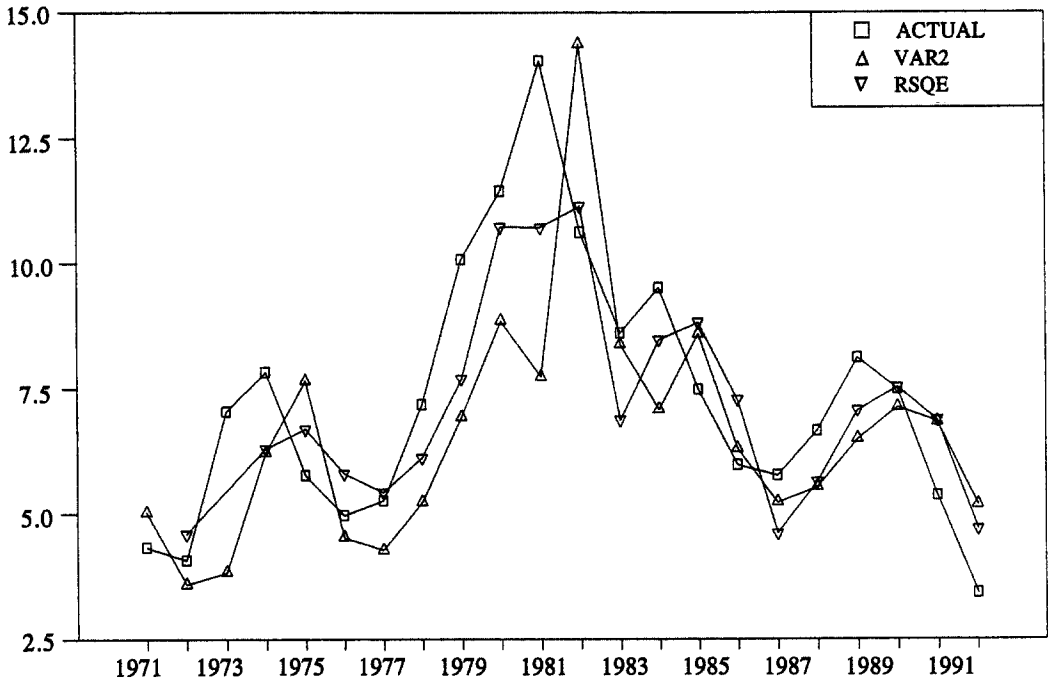
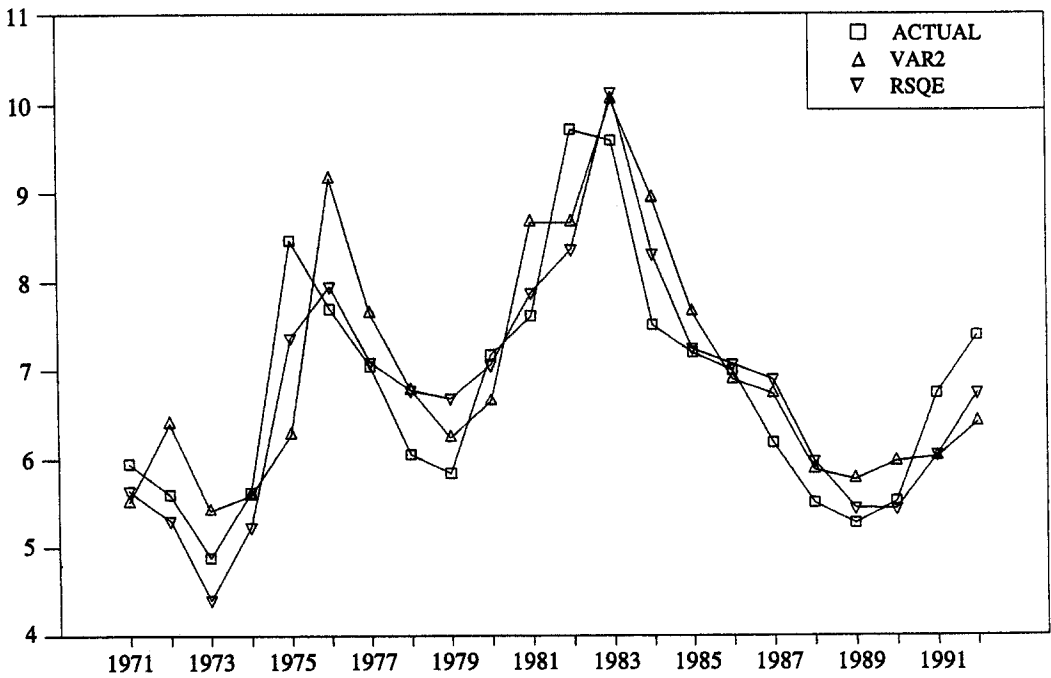


FIGURE IV
CIVILIAN UNEMPLOYMENT RATE



The similarity of the RSQE and VAR2 forecasts for all four variables is readily apparent. With the possible exception of the Treasury bill rate, it is difficult to distinguish between the RSQE and VAR2 forecasts on the basis of simple visual inspection of the graphs.

The root mean squared errors of the RSQE and VAR forecasts, displayed in Table 2, are potentially more informative.⁴ First, confining attention to the VAR forecasts, the forecast error⁵ statistics indicate that the second order VAR model produces the best forecasts for all four variables. Note that this is consistent with the BIC-estimated order of the VAR model. As the order of the VAR model is increased beyond two, there is a noticeable decline in the accuracy of the forecasts. This result illustrates the potentially deleterious effect of overparameterization of a VAR model.

TABLE 2

Root Mean Squared Forecast Errors: 1971-92

Forecast	Variable			
	<i>Y</i>	<i>P</i>	<i>R</i>	<i>U</i>
RSQE	1.26	1.40	1.37	0.59
VAR1	1.27	1.88	2.24	0.90
VAR2	1.25	1.55	2.23	0.86
VAR3	1.37	1.77	2.29	0.89
VAR4	1.90	2.11	2.47	0.92
VAR5	1.88	1.98	2.46	0.92
VAR6	2.14	1.94	2.59	0.98
VAR7	2.22	1.95	2.70	1.07
VAR8	2.71	2.23	2.95	1.16

Notes: *Y* = annual rate of growth of real GNP (%).

P = annual rate of inflation of the GNP Deflator (%).

R = annual average Treasury Bill Rate.

U = annual average civilian unemployment rate.

⁴ The use of root mean squared errors to compare models is standard in the literature, but it is not without its critics. See, for example, Clements and Hendry [1993].

⁵ Forecast errors were obtained by subtracting predicted values from the actual values as reported in the December 1994 NIPA release. Thus the most recently released data are treated as the "true" values. Implicitly, it is assumed that the purpose of forecasting is to predict the true value of the variable, which may not be known for many years after the fact, not the first-release value.

The RSQE forecasts are better on average than the best of the VAR forecasts for three of the four variables. The sole exception is the rate of growth of real GNP, for which the VAR2 model produces forecasts with a slightly smaller root mean squared error.

Root mean squared error is a function of the variance and the bias of the forecast error. Estimates of forecast bias, defined as the expectation of actual-minus-predicted values, are shown in Table 3. It is interesting to note that the VAR model forecasts are remarkably free of forecast bias. Over this period, RSQE forecasts of the real growth rate exhibit a small but statistically significant upward bias, and the inflation rate forecasts exhibit a small but statistically significant downward bias.

TABLE 3

Estimates of Forecast Bias: 1971-92

Forecast	Variable			
	<i>Y</i>	<i>P</i>	<i>R</i>	<i>U</i>
RSQE	-0.54 (.04)	0.58 (.05)	0.33 (.29)	0.02 (.87)
VAR1	-0.09 (.75)	-0.64 (.11)	0.30 (.54)	-0.17 (.39)
VAR2	-0.12 (.65)	-0.09 (.79)	0.73 (.13)	-0.17 (.36)
VAR3	0.03 (.93)	-0.13 (.74)	0.58 (.24)	-0.21 (.27)
VAR4	0.27 (.51)	-0.16 (.73)	0.48 (.37)	-0.33 (.10)
VAR6	-0.67 (.15)	-0.20 (.63)	0.62 (.27)	-0.41 (.04)
VAR8	-1.26 (.03)	-0.12 (.81)	0.50 (.44)	-0.26 (.31)

Note: Two-sided *p* values are shown in parentheses.

With the exception of the Treasury bill rate, the differences in root mean squared errors of the RSQE and VAR2 forecasts are small. A regression procedure described by Williams [1959] can be used to test the hypothesis that the forecast error variances of two alternative forecast procedures are equal, with the observed differences due to chance variations in the data. Let f_1 and f_2 denote alternative forecasts of the variable y . The procedure involves a regression through the origin of $y - [(f_1 + f_2)/2]$ on $f_2 - f_1$. A

significant positive coefficient indicates that f_2 predicts better than f_1 . A significant negative coefficient indicates that f_1 predicts better than f_2 . The t statistics for a test of the equal variance hypothesis are shown in Table 4. The level of significance at which the null hypothesis of equal variances would be rejected is shown in parentheses. These results show that for both the Treasury bill rate and the unemployment rate, the equal-variance hypothesis is uniformly rejected using the usual 5 or 10 percent level of significance. There is such a high degree of collinearity of the RSQE and low-order VAR forecasts of output growth and inflation that it is not possible to reject the hypothesis that the observed differences in the root mean squared errors are due to chance variations in the data.⁶

TABLE 4
Test Statistics for Equality of
RSQE and VAR Forecast Error Variances

Forecast	Variable			
	<i>Y</i>	<i>P</i>	<i>R</i>	<i>U</i>
VAR1	-0.02 (.98)	-1.65 (.11)	-4.08 (.00)	-7.22 (.00)
VAR2	0.04 (.97)	-0.72 (.48)	-3.94 (.00)	-6.09 (.00)
VAR3	-0.42 (.68)	-1.60 (.12)	-3.80 (.00)	-7.17 (.00)
VAR4	-2.03 (.06)	-2.92 (.01)	-3.60 (.00)	-8.05 (.00)
VAR6	-3.04 (.01)	-2.24 (.04)	-3.63 (.00)	-5.61 (.00)
VAR8	-5.00 (.00)	-3.29 (.00)	-3.96 (.00)	-5.93 (.00)

Notes: Two-sided p values are shown in parentheses. (.00) indicates a values less than .01.

⁶ These results are broadly consistent with the results reported by McNees [1986] for a comparison of Litterman's [1986] Bayesian VAR (BVAR) model with a number of macroeconomic forecasts, including RSQE forecasts. McNees found that for the period 1980:2 - 1985:1, quarterly BVAR forecasts were among the most accurate for real GNP and the unemployment rate, the least accurate for the implicit price deflator, and among the least accurate for the Treasury bill rate.

IV. Forecast Encompassing Tests

Turning now to the question of the conditional efficiency of RSQE forecasts, in order to determine if RSQE forecasts are efficient, one asks if they could be improved by combining them with VAR forecasts.⁷ For each variable, the actual value is regressed on the RSQE forecast and the VAR forecast. A statistically significant coefficient for the RSQE forecast indicates that the RSQE forecast contains information that is not in the VAR forecast. Similarly, a significant coefficient for the VAR forecast indicates that the VAR forecast contains information that is not contained in the RSQE forecast.

The results are shown in Tables 5-8 for each of the four variables. Following Fair and Shiller [1990], the *p* values shown in these tables are based on robust estimates of standard errors obtained using White's [1980] method. The first equation in each of these tables represents the baseline case in which the actual value of the variable is regressed on the RSQE forecast only. The standard error (SE) of these regressions provides a basis for comparison with the other regressions in the table.

TABLE 5

Univariate Encompassing Regressions: Real GNP

	Constant	RSQE	VAR	R ²	SE	DW
RSQE	-0.535 [.47]	1.000 [.10]		0.749	1.199	1.870
VAR1	-1.355 (.01)	0.541 (.00)	0.849 (.00)	0.854	0.937	1.940
VAR2	-0.784 (.10)	0.620 (.00)	0.532 (.01)	0.810	1.070	2.490
VAR3	-0.694 (.15)	0.707 (.00)	0.421 (.07)	0.794	1.114	2.220
VAR4	-0.522 (.25)	0.883 (.00)	0.153 (.36)	0.762	1.198	2.110
VAR6	-0.751 (.25)	0.961 (.00)	0.105 (.59)	0.753	1.222	2.060
VAR8	-0.287 (.52)	1.028 (.00)	-0.088 (.52)	0.752	1.222	1.710

Notes: Robust estimates of standard errors are shown in brackets. Two-sided *p* values based on robust estimates of the standard errors are shown in parentheses.

⁷ For a general discussion of the rationale for combining forecasts and various methods for combining forecasts, see Chapter 9 of Granger and Newbold [1986]. For specific applications to macroeconomic forecasting models, see Nelson [1984], Lupoletti and Webb [1986], and Fair and Shiller [1989, 1990].

The results shown in Table 5 for GNP indicate that both the RSQE and the low-order (≤ 3) VAR forecasts are significant. Neither forecast alone is efficient: the RSQE forecasts contain information that is absent from the VAR forecasts and the VAR forecasts contain information that is absent from the RSQE forecasts. Interestingly, the forecasts of higher-order (≥ 4) VAR models are not significant. If only higher-order VAR model forecasts had been examined, the results would have shown that RSQE forecasts of real GNP growth were efficient, according to this definition of efficiency. This indicates a potentially significant sensitivity of the results of this type of test to the order of the time series model that is used in the test. In this particular application, however, the results indicate that the inefficiency of RSQE forecasts is not great in the sense that the SE of the combined forecast is only slightly smaller than the SE of the RSQE forecast.

The results for the GNP deflator shown in Table 6 indicate that RSQE forecasts are efficient relative to low-order (≤ 3) VAR forecasts, but are not efficient relative to higher-order VAR forecasts. Despite the statistical significance of the higher-order VAR forecasts, there is little gain in forecast accuracy as measured by the SE of the combined-forecast regression. It should be noted that the Durbin-Watson statistic is on the low side. When the regression equations are corrected for first-order serial correlation using the Cochrane-Orcutt procedure, the statistical significance of the VAR6 and VAR8 forecasts persists, however.

TABLE 6

Univariate Encompassing Regressions: GNP Deflator

	Constant	RSQE	VAR	R ²	SE	DW
RSQE	1.240 [.58]	0.874 [.10]		0.709	1.300	0.690
VAR1	1.276 (.05)	0.969 (.01)	-0.083 (.81)	0.711	1.331	0.670
VAR2	1.240 (.04)	0.876 (.03)	-0.002 (.99)	0.709	1.334	0.690
VAR3	1.244 (.03)	1.134 (.00)	-0.230 (.50)	0.718	1.314	0.710
VAR4	1.037 (.04)	1.435 (.00)	-0.457 (.00)	0.756	1.222	0.960
VAR6	1.456 (.02)	1.281 (.00)	-0.390 (.13)	0.743	1.255	0.740
VAR8	1.786 (.00)	1.294 (.00)	-0.463 (.01)	0.781	1.158	0.830

Notes: Robust estimates of standard errors are shown in brackets. Two-sided p values based on robust estimates of the standard errors are shown in parentheses.

Table 7 shows that the RSQE forecasts of the Treasury Bill Rate uniformly dominate the VAR forecasts. The results in Table 8 show that the VAR forecasts of the unemployment rate contain information that is not captured by the RSQE forecast. Again, the improvement in forecast accuracy of the combined forecast is not very impressive, but it is statistically significant.

TABLE 7

Univariate Encompassing Regressions: Treasury Bill Rate

	Constant	RSQE	VAR	R ²	SE	DW
RSQE	-0.808 [1.11]	1.159 [.16]		0.748	1.367	1.980
VAR1	-1.176 (.44)	1.140 (.00)	0.071 (.39)	0.765	1.330	1.910
VAR2	-1.175 (.45)	1.136 (.00)	0.075 (.70)	0.765	1.329	1.910
VAR3	-1.050 (.47)	1.137 (.00)	0.055 (.77)	0.764	1.331	1.920
VAR4	-1.099 (.45)	1.132 (.00)	0.066 (.73)	0.765	1.331	1.920
VAR6	-1.033 (.49)	1.138 (.00)	0.050 (.79)	0.764	1.331	1.940
VAR8	-0.915 (.48)	1.147 (.00)	0.026 (.86)	0.764	1.332	1.950

Notes: Robust estimates of standard errors are shown in brackets. Two-sided p values based on robust estimates of the standard errors are shown in parentheses.

The previous tests have involved univariate-forecast-encompassing tests in the sense that they have involved the combination of forecasts of only a single variable from two different models. Multivariate-encompassing tests are potentially even more informative. In what follows, two types of multivariate-forecast-encompassing tests are examined. In the first case, the internal consistency of a vector of forecasts is tested in the following way: the actual value of the variable is regressed on the predicted value of that variable and the predicted values of a subset of the other variables in the model. If the forecast is internally consistent, the coefficients of the other predicted values would be expected to be zero, or at least not significantly different from zero. If this condition is satisfied, the forecast is said to be internally consistent; otherwise, the forecast is said to be internally

inconsistent. The test statistic that is used is the standard Wald-type test statistic denoted in the table by the symbol χ^2 .

TABLE 8

Univariate-Encompassing Regressions: Unemployment Rate

	Constant	RSQE	VAR	R ²	SE	DW
RSQE	0.667 [.51]	0.905 [.08]		0.802	0.601	1.580
VAR1	0.543 (.19)	0.779 (.00)	0.139 (.00)	0.857	0.524	1.140
VAR2	0.602 (.12)	0.765 (.00)	0.153 (.00)	0.855	0.529	1.020
VAR3	0.774 (.04)	0.759 (.00)	0.131 (.00)	0.851	0.536	1.090
VAR4	0.930 (.01)	0.737 (.00)	0.128 (.00)	0.856	0.527	1.130
VAR6	1.053 (.01)	0.691 (.00)	0.158 (.02)	0.847	0.543	1.300
VAR8	1.295 (.01)	0.674 (.00)	0.138 (.14)	0.827	0.576	1.490

Notes: Robust estimates of standard errors are shown in brackets. Two-sided p values based on robust estimates of the standard errors are shown in parentheses.

The results of the internal consistency tests for RSQE and VAR2 forecasts are shown in Table 9. The results shown in this table are largely negative. The only case for which the forecast consistency condition is satisfied is the RSQE forecast of the interest rate. In all other cases, same-model forecasts of other variables contain information about the forecast variable. For the case of RSQE forecasts of GNP, the result is not very informative for diagnostic purposes, since none of the individual other forecast coefficients are individually significant, but as a set they are significant. An examination of the pattern of coefficients in these regressions does not appear to lead to any particularly useful diagnostic results, with the possible exception of the RSQE forecasts of the rate of price inflation. The large negative coefficient on the predicted unemployment rate seems to indicate that the short-run tradeoff between inflation and unemployment exhibited by the data is not captured very well by the model.

A second type of multivariate-forecast-encompassing test involves testing the conditional efficiency of the forecasts. The test regression is similar to the consistency test regression, except that the other variables in the regression are the forecasts from the other model. If the coefficients of the forecasts of the variables from the other model are not significant, the forecast is said to be efficient.

TABLE 9
Multivariate Tests of Internal Consistency of
RSQE and VAR2 Forecasts: 1971-92

Dependent Variable	RSQE Forecasts			VAR2 Forecasts			Constant	R ²	SE	DW	χ^2	
	Y	P	R	U	Y	P						R
Y	1.026 (.12)	0.130 (.14)	-0.300 (.14)	0.200 (.20)	0.562 (.18)	-0.140 (.12)	0.616 (.17)	-0.442 (.10)	0.784	1.233	1.770	10.520 [.01]
P	0.340 (.08)	1.386 (.07)	-0.184 (.08)	-0.838 (.14)	-0.046 (.19)	1.007 (.12)	-0.222 (.18)	-0.518 (.13)	0.938	0.690	2.400	43.580 [.00]
R	0.240 (.20)	0.352 (.20)	1.033 (.19)	-0.513 (.31)	0.266 (.46)	0.434 (.21)	0.081 (.46)	0.516 (.38)	0.787	1.377	1.940	3.380 [.34]
U	-0.188 (.07)	-0.030 (.07)	0.004 (.08)	0.968 (.10)	-0.226 (.13)	-0.016 (.05)	0.739 (.10)	0.102 (.09)	0.500	2.050	1.580	17.400 [.00]
									0.838	0.586	1.290	8.680 [.03]
									0.800	0.655	1.680	74.250 [.00]

Notes: Robust estimated standard errors are shown in parentheses. Significance levels are shown in brackets. An entry [.00] indicates a significance level less than .01. χ^2 is the χ^2 statistic for a Wald-type test of the hypothesis that the coefficients of the forecasts other than the variable being forecast are zero.

TABLE 10

Multivariate Tests of Efficiency of
RSQE and VAR2 Forecasts: 1971-92

Dependent Variable	RSQE Forecasts				VAR2 Forecasts				Constant	R ²	SE	DW	χ ²	
	Y	P	R	U	Y	P	R	U						
Y	0.487 (.12)	0.168 (.14)	-0.198 (.11)	0.123 (.22)	0.288 (.16)	-0.036 (.12)	0.417 (.17)	-0.372 (.09)	0.023 (1.19)	0.862	0.994	1.720	29.420 [.00]	
	0.742 (.15)	0.168 (.14)	-0.198 (.11)	0.123 (.22)	0.465 (.18)	-0.036 (.12)	0.417 (.17)	-0.372 (.09)	-1.288 (1.64)	0.825	1.149	2.560	29.800 [.00]	
P	0.931 (.26)	1.585 (.19)	-0.159 (.06)	-0.888 (.15)	0.118 (.13)	0.261 (.25)	-0.459 (.13)	-0.361 (.09)	4.670 (1.00)	0.902	0.844	1.260	102.410 [.00]	
	0.345 (.07)	1.585 (.19)	-0.159 (.06)	-0.888 (.15)	0.118 (.13)	0.261 (.25)	-0.459 (.13)	-0.361 (.09)	4.658 (1.05)	0.941	0.697	2.480	92.820 [.00]	
R	0.447 (.28)	0.403 (.18)	1.105 (.19)	0.386 (.84)	0.006 (.27)	0.112 (.17)	0.101 (.33)	-0.323 (.22)	-1.399 (1.64)	0.779	1.451	1.800	2.870 [.58]	
	0.447 (.28)	0.403 (.18)	1.105 (.19)	0.386 (.84)	0.006 (.27)	0.112 (.17)	-0.653 (.97)	-0.323 (.22)	-1.911 (1.55)	0.792	1.407	2.080	74.300 [.00]	
U	-0.072 (.08)	0.044 (.08)	-0.217 (.07)	0.822 (.10)	0.757 (.22)	-0.189 (.09)	-0.016 (.05)	0.102 (.22)	0.048 (.08)	1.241 (.55)	0.881	0.521	1.410	11.490 [.02]
	-0.072 (.08)	0.044 (.08)	-0.217 (.07)	0.822 (.10)	0.757 (.22)	-0.189 (.09)	-0.016 (.05)	0.102 (.22)	0.249 (.06)	0.986 (.55)	0.881	0.520	1.150	131.590 [.00]

Notes: Robust estimated standard errors are shown in parentheses. Significance levels are shown in brackets. An entry [.00] indicates a significance level less than .01. χ^2 is the χ^2 statistic for a Wald-type test of the hypothesis that the coefficients of the other model forecasts are zero.

The results of the efficiency tests are shown in Table 10. The χ^2 statistics and significance levels for testing consistency are given in the last column of the table. Not surprisingly, given the univariate test results in Tables 5-8, the multivariate test results indicate that except for the RSQE interest rate forecasts, neither the RSQE nor the VAR forecasts are efficient. Statistically significant, but again limited, gains in forecast accuracy could be obtained by combining forecasts from the different models. One of the interesting conclusions that emerges from this table is that improvements in forecast accuracy are not confined to a univariate combination of forecasts. For example, in the regression of the real GNP growth rate on the RSQE forecast and the VAR2 forecasts of Y , P , R , and U , the coefficients of R and U are both significantly different from zero, while the coefficient of Y is marginally significant.

The regression results shown in Tables 9 and 10 indicate that there is a statistical basis for concluding that neither RSQE nor VAR forecasts are conditionally efficient. However, as Nelson [1972], Diebold [1989], and Clements and Hendry [1993] have pointed out, combining forecasts is not a theoretically attractive way to try to improve forecast accuracy. A more attractive procedure, at least in principle, is to combine the information sets on which the forecasts are based. The forecast-encompassing tests indicate that there may be some benefit in doing this.

V. Conclusion

The results of this investigation can be summarized as follows:

1) Based on root mean squared error comparisons for the period 1971-92, the RSQE forecasts are more accurate than VAR forecasts for annual averages of the rate of inflation, the interest rate, and the unemployment rate. The VAR2 forecast is more accurate than the RSQE forecast of the average annual rate of growth of real GNP. Using a standard test, the differences in root mean squared errors are not statistically significant for the forecasts of the real GNP growth rate and the rate of inflation. The differences are statistically significant for interest rate and unemployment rate forecasts.

2) Univariate and multivariate encompassing tests indicate that, except for the RSQE interest rate forecasts, neither RSQE nor VAR forecasts are conditionally efficient.

3) There is some indication that an improved forecast could be obtained by combining RSQE and VAR forecasts. However, the estimated standard errors of the forecast-encompassing regressions indicate that the potential improvement in forecast accuracy is rather limited.

The empirical results in this paper underscore several points that should be emphasized in connection with the comparison of VAR and econometric model forecasts. First, care needs to be exercised to avoid overparameterizing the benchmark VAR model. If too many lags are included in the model, outside-sample-forecast accuracy may be reduced and forecast-encompassing tests may not detect forecast inefficiency. Model selection criteria can help to avoid the problem of overparameterization associated with using too many lagged values. A similar difficulty may be associated with including too many variables in the VAR model, but this has not been investigated in this paper. Second, root mean squared error comparisons may not reveal model inadequacy. It is common to find that VAR models generate forecasts that rival econometric model forecasts in terms of root mean squared error accuracy. Forecast-encompassing tests may reveal inadequacies that are not apparent from mean squared error comparisons. Finally, multivariate-

encompassing tests are a potentially richer source of information about model adequacy than are univariate-encompassing tests.

REFERENCES

- Clements, Michael P.; Hendry, David F. "On the Limitations of Comparing Mean Square Forecast Errors," *Journal of Forecasting*, 12, 1993, pp. 617-37.
- Diebold, Francis X. "Forecast Combination and Encompassing: Reconciling Two Divergent Literatures," *International Journal of Forecasting*, 5, 1989, pp. 589-92.
- Donihue, Michael R. "Evaluating the Role Judgement Plays in Forecast Accuracy," *Journal of Forecasting*, 12, 1993, pp. 81-92.
- Fair, Ray C.; Shiller, Robert J. "Comparing Information in Forecasts from Econometric Models," *American Economic Review*, 80, 1990, pp. 375-89.
- _____. "The Informational Content of Ex Ante Forecasts," *Review of Economics and Statistics*, 71, 1989, pp. 325-31.
- Granger, C. W. J.; Newbold, Paul. *Forecasting Economic Time Series*, 2nd ed., New York: Academic Press, 1986.
- Green, R. Jeffery; Hickman, Bert G.; Howrey, E. Philip; Hymans, Saul H.; Donihue, Michael R. "The IS-LM Core of Three Econometric Models," in Lawrence R. Klein, ed., *Comparative Performance of U.S. Econometric Models*, New York: Oxford University Press, 1991, pp. 89-124.
- Hottelling, Harold; Working, Holbrook. "Applications of the Theory of Error to the Interpretation of Trends," *Journal of the American Statistical Association, Papers and Proceedings*, 24, 1929, pp. 73-85.
- Hottelling, Harold. "The Generalization of Student's Ratio," *Annals of Mathematical Statistics*, 21, 1931, pp. 360-78.
- _____. "The Selection of Variates for Use in Prediction With Some Comments on the General Problem of Nuisance Parameters," *Annals of Mathematical Statistics*, 11, 1940, pp. 271-83.
- _____. "Problems of Prediction," *American Journal of Sociology*, 48, 1942, pp. 61-76.
- Howrey, E. Philip; Hymans, Saul H. "An Open Economy Analysis of the Dynamic Properties of the Michigan Quarterly Econometric Model of the U.S. Economy," in M. Dutta, ed., *Economics, Econometrics and the Link: Essays in Honor of Lawrence R. Klein*, Amsterdam: North Holland, 1995, pp. 129-58.
- Klein, Lawrence R.; Goldberger, Arthur S. *An Econometric Model of the United States Economy: 1929-1952*, Amsterdam: North Holland, 1955.
- Litterman, Robert B. "Forecasting With Bayesian Vector Autoregressions - Five Years of Experience," *Journal of Business and Economic Statistics*, 4, 1986, pp. 25-38.
- Lucas, Robert E. Jr.; Sargent, Thomas J. "After Keynesian Macroeconomics," in Robert E. Lucas Jr.; Thomas J. Sargent, eds., *Rational Expectations and Econometric Practice*, Minneapolis: University of Minnesota Press, 1981.
- Lupoletti, William M.; Webb, Roy H. "Defining and Improving the Accuracy of Macroeconomic Forecasts: Contributions from a VAR Model," *Journal of Business*, 59, 1986, pp. 263-85.
- Lütkepohl, Helmut. "Comparison of Criteria for Estimating the Order of a Vector Autoregressive Process," *Journal of Time Series Analysis*, 6, 1985, pp. 35-52.
- McNees, Stephen K. "Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts," *Journal of Business and Economic Statistics*, 4, 1986, pp. 5-15.
- Nelson, Charles R. "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy," *American Economic Review*, 62, 1972, pp. 902-17.
- _____. "A Benchmark for the Accuracy of Econometric Forecasts of GNP," *Business Economics*, 19, 1984, pp. 52-8.
- Sims, Christopher A. "Macroeconomics and Reality," *Econometrica* 48, 1980, pp. 1-48.
- White, Halbert. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 1980, pp. 817-38.
- Williams, E. J. *Regression Analysis*, New York: Wiley, 1959.