Has Macro-Forecasting Failed?

Victor Zarnowitz

What to Ask, Why, and How?

The title of my paper, "Has Macro-Forecasting Failed?" serves a good purpose, even though it seems to be somewhat provocatively phrased. We are reminded that claims to predict the future must be properly modest or they will prove disappointing. The more that forecasts matter and the more people that depend on them, the greater the dangers of having overstated promises and unrealistically high expectations—and macroeconomic forecasts matter greatly when used as guides by public and private decisionmakers.

The question deserves a straightforward, but careful, answer. A simple "yes" or "no" lacks meaning. Forecasting the economy's course, even short-term and in the broadest outline, is a mixture of art and science that can be very imperfect and sophisticated at the same time. Thus, we face a problem whose solution depends on the treatment of more fundamental questions about (1) what the forecasts are and why they are needed, and (2) what we can reasonably expect of them.

It is easy to think of needs, uses, and standards associated with macro-forecasting that will readily show it as a failing enterprise. What is more difficult but also more important is to decide which legitimate and credible applications of forecasting would, in principle, allow our title question to be interesting (i.e., capable of being answered either positively or negatively according to some sensible and, so far as possible, quantifiable criteria).

To establish what forecasters can and should do, we must study the record and assess the probable future of their endeavors. I can sum up these large subjects only selectively in this paper. Thus, I

Cato Journal, Vol. 12, No. 1 (Spring/Summer 1992).

The author is Professor Emeritus of Economics and Finance at the University of Chicago Graduate School of Business and a Research Associate of the National Bureau of Economic Research.

shall concentrate on what I studied firsthand, namely annual and multiperiod quarterly forecasts of the principal aggregate variables in the United States during the post–World War II era. The predictions refer to levels and changes of national output, employment and unemployment, consumption and investment, prices, and interest rates. They cover short and intermediate horizons of one or two years along with one to four, or at most eight, quarters ahead.

Movements associated with the business cycles prevail over such time spans. Irregular variations from random causes and intrayear, approximately periodic, seasonal movements influence heavily most economic time series in the shortest run measured in weeks and months. But the forecasts are generally in quarterly and annual units, and they aim at seasonally adjusted values wherever seasonal movements exist. Furthermore, the random noise in the series is unforecastable. Hence, the systematic part of the time series covered by the macro-forecasts consists mainly of cyclical movements and, to a lesser extent, elements of longer trends. It follows that a forecast should be judged successful if it approximates reasonably well that part of its target. The trend-cycle movements include the effects of past shocks and seasonal innovations that may or may not be knowable, but those movements presumably have important endogenous ingredients as well.

The task of forecasting is more difficult than the term "systematic" may suggest. Business cycles are persistent and recurrent, but they are by no means predetermined or periodic. They tend to be pervasive but affect different variables and sectors in different ways. Fluctuations and long trends in real growth and inflation interact with each other and contain stochastic elements. The economy in motion is a complex of dynamic processes, subject not only to a variety of disturbances but also to gradual and discrete changes in structure, institutions, and policy regimes. No wonder there are few, if any, constant quantitative rules (e.g., time-invariant linear econometric equations) to help the macro-forecaster effectively and consistently over more than a few years or from one business cycle to another.

Indeed, in social sciences and human affairs generally, it is only prudent to recognize from the outset that the future simply cannot be foreknown. Any maker or user of economic forecasts, therefore, must always be prepared to be wrong, even if the errors are, at best, relatively small and unpredictable. However, this probability of error does not alter the fact that most decisions that matter are inevitably forward-looking and, hence, involve forecasts. Where macroeconomic forecasting cannot be avoided, it is probably advisable to make those forecasts explicit and as good as possible, given the available cost-efficient methods and information. Forecasts are needed and can be useful even when imprecise.

My basic approach, therefore, is to ask the following of the available data: What professional standards have economists who are engaged in macro-forecasting been able to attain and maintain in competing with each other and alternative methods? We have learned much in recent times by assembling and examining measures of absolute and relative accuracy for reasonably representative samples of macroeconomic forecasts (my own published work in this area goes back to 1967). By now we would expect most professional predictions to be, on average and over time, much better than naive mechanical projections, but that expectation is a minimal requirement: To be successful, predictions should exceed significantly the more sophisticated univariate and multivariate time-series models. Further, it is desirable that the forecasts be free of such systematic errors as could have been either prevented by good modeling or eliminated by learning from the past. The extent of such biases depends on the stability of economic processes, the lags and costs of adjustments to unanticipated change, and the relative contributions to the forecasts made by models and techniques on the one hand and by new informational and judgmental inputs on the other.

In principle, what matters most about macroeconomic forecasts is their usefulness for decisionmakers both in government and in the private sector. True, this criterion is most difficult to apply directly as little is known generally about the loss functions of users and the effective costs and returns to them associated with such forecasts. But even here not all is lost; we can safely assume that a high positive correlation exists between the usefulness of forecasts and their measurable quality attributes, notably high relative accuracy.

Forecasting as a Competitive Industry: Why No One Is Best

Peering into the future is an ancient occupation frequently characterized by great hazards and by corresponding vagueness or obfuscation. Economic prediction as an artful pursuit or a game of chance undoubtedly has a long past, but authentic forecasting of well-specified future values of aggregative variables is of recent origin. A responsible appraisal requires a recorded history of forecasts that are not only explicit and verifiable but also sufficiently numerous and consistent. This requirement rules out vast amounts of data, both old and new, and urges concentration on the longest available time series of reasonably comparable predictions from reputable sources.

Economic and financial forecasting in the United States today is an industry of significant size. Many forecasters belong to the National Association of Business Economists, whose membership numbered more than 3,300 in 1990 (it numbered 322 in 1959, the year NABE was founded). The forecasting units vary from individuals and small teams to sizable divisions of some large corporations and multibranch specialized consulting firms. Some of the latter operate large-scale econometric models and provide customer services internationally. Business demand for forecasts of numerous more or less aggregative variables is largely satisfied by subscriptions to such services. For small numbers of primary macrovariables, special publications survey groups of professional forecasters monthly or quarterly. Thus, the U.S. market for these forecasts can probably be described as a mixture of competitive and oligopolistic elements, with the overall number of sellers relatively large and the barriers to entry low.[1] In addition, some macro-forecasts are provided by government agencies, essentially as public goods, and some can be acquired at very low cost from the press. U.S. government forecasts are designed to serve as inputs into economic policymaking, and they originate in several agencies. Some forecasts are publicized but most are for internal uses only.

Forecasters compete, adapt to continuous change and new developments in the economy, and try to improve and differentiate their products. Few leave their models and techniques unchanged for long. Moreover, success in forecasting may be occasional and fortuitous or intuitive. Hence, a particular forecaster's record may not be reliable as a basis for inferences on how he or she will perform in the future. The shorter that record, the more uncertain are such inferences. Ranking the forecasters on how well they predicted changes in a single short period is quite risky and not very informative; in the next period the ranks are very likely to differ considerably. Nevertheless, such comparisons are commonly made at least once each year in the business press.

That there is great inerest in finding out "Who forecasts best?" is certainly not surprising, and many recent studies address this question. However, their well-established general conclusion is that no one forecaster does or, equivalently, that many forecasters do. That is, the measures of overall accuracy surveyed (typically, mean absolute errors [MAEs] or root mean square errors [RMSEs]) do not show

[1]In other highly developed countries, macroeconomic forecasting is generally more concentrated in a few private sources or in government agencies and publicly supported research organizations.

any of the compared individuals or organizations to be consistently and generally superior to others. The rankings vary for different periods, variables, and horizons covered; they also depend on the criteria and measures applied. Moreover, the differences between the MAEs or RMSEs across the best-known sources are mostly small and of uncertain statistical significance (which cannot be directly tested since the forecasts are not independent).[2]

There are good probable reasons why the principal macro-forecasters cannot be ranked unambiguously by any standards of accuracy. Authors of predictions that are matched by variable, time of issue, and target period draw on much the same body of data; often use similar methods; are exposed to common current events and attitudes; and, to some extent, interact and influence each other directly. A free market exists in economic data and ideas, and advances in forecasting technology are soon open to all practitioners. This openness tends to reduce the diversity of individual predictions created by the undoubted fact that forecasters differ greatly in theoretical orientation and training and in talent and experience. It also presumably keeps the comparative advantages of the initially better-endowed forecasters more temporary and smaller than they would otherwise be.

Successful forecasting has diverse ingredients that are unlikely to be monopolized as a combination. Some functions, like the exploitation of time-series properties in the data, are best performed by the computer and its programs. Other functions, like the quick and efficient sifting of new information, would seem to require much specialized skill and experience. The ability to develop superior judgmental forecasts on this basis is probably a rare individual gift that cannot be easily taught, transferred to others, or applied on a large scale.

All of this helps to explain why concurrent matched predictions from different sources show both common trends and much dispersion around them. The frequently alleged predominance of a single

[2]For example, of the six sets of forecasts examined in Zarnowitz (1979), each was "best" for at least one variable, subperiod, and span considered. The comparative advantages were generally quite scattered, however, except that forecasts released later in a quarter, being based on more information, tended to be more accurate than those made earlier in the quarter. (This factor can be isolated by comparing early and late predictions from sources that forecast monthly or twice per quarter.) Similar results are reported elsewhere (e.g., in McNees 1979 and, for the United Kingdom, in Wallis 1989).

"consensus forecast" is mostly a myth.[3] The above arguments are also consistent with the observation that the interforecast differences are not persistent enough to give rise to systematic ratings of predictive performance.

The Multiplicity of Methods and Models

The coexistence of many different forecasts aiming at the same targets and continuing to have significantly dispersed errors is in itself an indication that no single model or technique is generally expected to prove consistently superior to others. If any such winner were believed to exist, it would soon come to enjoy the first preference of the profession and be universally adapted. Instead, the market has room for a sizable and diversified activity of macroforecasting.

Indeed, substantial and uncontroverted evidence from surveys of professional forecasters shows that they distinctly favor several approaches in varying combinations. For many years, questionnaires of the quarterly *Economic Outlook Survey*, jointly conducted by the American Statistical Association (ASA) and the National Bureau of Economic Research (NBER) during 1968–90, collected information on the premises and procedures incorporated in the members' forecasts. More than 70 percent reported using—and from half to two-thirds ranked first—the "informal GNP model" in which the major expenditure components of GNP are predicted in various ways, combined into an overall forecast, and then checked and adjusted for plausibility and internal consistency. This model is itself a mix of procedures applied eclectically and flexibly with large elements of judgment, not a well-defined method. Leading indicators were ranked second by most respondents and were used by large numbers. Somewhat smaller proportions of the survey membership mentioned anticipations surveys, which generally ranked lower. The percentage of users of outside econometric models rose from less to more than half between the late 1960s and early 1980s. About one-fourth of the respondents had their own econometric models, and the proportion of those ranking such models first was similar. Finally, other methods such as time-series models were specified by fewer than 20 percent

[3]A forecaster must indeed expect his results to be compared with that standard that is widely reported by professional associations, business magazines, and other media. But there are as many average forecasts as there are groups surveyed, and they may at times express little agreement. The averages are not well specified, and they lag somewhat behind the release of many noteworthy forecasts for the same period.

of the ASA/NBER survey participants and were preferred by about half of those respondents.

Different methods tend to complement each other (e.g., new readings on monthly cyclical indicators and the latest results from an investment or consumer expectations survey may modify forecasts from econometric models or the informal approach). This is the presumed reason why the dominant forecasting practice is to use various combinations of these techniques. Other sources confirm this important lesson.[4]

The forecasters' methodological choices, as reported in the ASA/NBER surveys, do not appear to be associated with significant differences in predictive accuracy. Those who preferred econometric models in their own work were not as a group systematically better than those who preferred the informal approach. However, according to comparisons based on first rank only, subscribers to outside models—a subset dominated by large companies using well-known econometric service bureaus and their own professional staffs—had a marginal advantage over the other categories. On the whole, these test results are consistent with the view that combining different procedures helps, particularly when done by experienced forecasters (Su and Su 1975; Zarnowitz 1971, 1984).

The 1950s and 1960s witnessed a great ascendancy of macroeconometric forecasting based largely on models of Keynesian provenance and the paradigm of "neoclassical synthesis." The prevailing belief was that business cycles could be subdued by fine-tuning fiscal and monetary policy, aided by sufficiently early and accurate predictions. The naively overconfident nature of such views became increasingly clear as forecasting and policy failures multiplied during the late 1960s and 1970s. Inflation and unemployment both trended upward, which contradicted the idea of a stable Phillips-curve tradeoff. New shocks to oil and other major input prices caused adverse shifts in aggregate supply, which undermined the effectiveness of aggregate demand management. Discretionary fiscal policies have long proved too sluggish and inflexible to be effectively countercyclical.

[4]Thus, the annual surveys of NABE in 1975–79 show that 52–60 percent of their members preferred "eclectic judgmental" and 22–28 percent "eclectic econometric" methods. A special mail survey sent to the Blue Chip forecasters in 1987 resulted in the following mix of average contributions: judgment, 48 percent; econometric model, 28 percent; and time-series analysis, 24 percent. Even the organizations with their own large-scale econometric models (e.g., BEA, Chase, DRI, Kent, UCLA, Wharton) assigned sizable weights to judgment (20–50 percent, on average 30 percent) and other elements such as time-series methods, current data analysis, and interaction with others (10–20 percent). See McNees (1981).

But discretionary monetary policies, though more potent, were frequently not any more timely and successful. This lack of success was not always or necessarily due to lacking foresight; often the fault lay in wrong indicators and targets, divided opinion and indecisiveness of authorities, or miscalculations of long and variable lag effects. The point is that monetarist models, too, failed to produce dependable forecasts for the conduct of macro-stabilization policies. The crudely monetarist-oriented Fed tactics that temporarily replaced the Keynesian regime in the late 1970s succeeded in finally eliminating the unbearably high and volatile inflation and interest rates, but only at much higher costs than expected and by contributing heavily to the two recessions of 1980–82.

Criticism of both Keynesian and monetarist model specifications and predictions naturally led to new developments, some of which were much needed and promising: a resurgence of theoretical and empirical studies on business cycles, on the formation and role of expectations, and on the microfoundations of aggregate supply and price adjustment. New theories emerged, stressing rational expectations, imperfect information, real business cycles, wage and price contracts, and rigidities. But the new ideas and methods are still in the stages of development and academic debate: They have not yet given rise to econometric models and demonstrated their usefulness for macro-forecasting.

People who must predict changes in the economy at frequent and regular intervals are typically absorbed by the technical requirements of monitoring and processing information, analyzing current developments, and preparing interpretive reports. Most are pragmatic in using any data and approaches deemed helpful; few spend much time on working with specific theoretical models. The task of testing various hypotheses, models, and methods is left largely to the academic economists. A very recent study of forecasters who were cross-classified according to their theoretical preferences as well as methodological choices concluded that "No one ideology or technique yields consistently more accurate forecasts than others."[5]

[5]See Batchelor and Dua (1990, p. 3). They used the Blue Chip 1987 survey mentioned in note 4 above and examined annual forecasts of real growth, inflation, and interest rates made in 1976–86 by 44 respondents. The weights placed on the listed theories were as follows: Keynesian 43 percent, monetarist 20 percent, supply side 12 percent, RE 8 percent, Austrian 4 percent, and other 13 percent. Some support was found for the inference that the Keynesian-econometric combination had an advantage, but this could reflect the fact that the modern versions of other theories and methods developed later and so had adherents with less practical and diverse experience. The forecasters in the sample generally used elements of more than one theory and relied on more than one technique.

How Accurate Have the Forecasts Been?

Progress in forecasting, as distinguished from occasional successes that may be due to chance, can come only from advances of science in discovering and quantifying important regularities. Although such predictive relationships are found in macroeconomics, their range and duration are probably more limited than they are widely believed to be. The economy grows and fluctuates in various ways, reflecting the diversity of human behavior that causes and reacts to the omnipresent change. Of necessity, economic theory simplifies starkly the motivations of individual and collective action, thereby attempting to reduce the uncertainty surrounding economic change. In the process, economists risk taking their models too seriously and overestimating their ability to predict the movements of the real economy. As already noted, some did succumb to this error in the recent past when the economy enjoyed relatively stable growth and rising prosperity, while macroeconomics also appeared to be doing well. Under such conditions, the informed lay opinion was only too willing to accept the optimistic claims of the experts. The present danger is one of overreaction in the opposite direction.

Actually, the true macro-forecasts from credible sources were never as good as many had once believed; nor were they as bad as some now claim. There is no evidence that forecasts have deteriorated over time; indeed, the opposite view—that some significant improvement has occurred—finds more support in the data, as will be shown directly. So the question "Has macro-forecasting failed?" cannot be answered in the affirmative simply on this ground.

Table 1 is based on the 1953–89 series of predicted annual rates of change in aggregate demand, output, and the price level. It covers a large number of forecasts from a variety of sources: business economists and others employed by private companies in manufacturing, finance, trade, consulting, etc.; academic and research institutions; and the president's Council of Economic Advisers (CEA), which prepares the principal government forecast. Predominantly judgmental predictions that originate mainly in business are represented along with predictions made by econometricians working with large models.[6] Although the included forecast sets differ in many respects, they are treated as sufficiently comparable for our present purpose, which is to consider the broad trends over time in the overall forecasting accuracy. The question of who forecasts best is of no interest here. All the predictions covered are made around the end of the

[6]Michigan and Wharton are the oldest families of such models in use. For a list of the coded sources, see Table 1, note a.

TABLE 1

ANNUAL FORECASTS OF PERCENTAGE CHANGES IN NOMINAL AND REAL GNP AND THE IMPLICIT PRICE DEFLATOR: SUMMARY MEASURES OF ERROR, 1953–89

Line	Period Covered (1)	(No. of Years)	Forecasts (Code)[a] (2)	Mean Abs. Errors (MAE)[b] Mean (3)	Mean Abs. Errors (MAE)[b] Range (4)	Mean Errors (ME)[c] Mean (5)	Mean Errors (ME)[c] Range (6)	Relative Error[d] (7)	Mean Abs. % Change[e] (8)
				Gross National Product (GNP)[f]					
1	1953–76	(24)	1, 2	1.4	.4	−.8	.3	.6	6.6
2	1956–63	(8)	1, 2, 4	1.6	.3	−.4	.7	.8	5.0
3	1963–76	(14)	1, 2, 6, 7, 8	1.0	.5	−.5	.6	.6	7.9
4	1969–76	(8)	1, 2, 5, 6, 7, 8	.9	.4	−.1	.6	.5	8.4
5	1969–89	(21)	5, 6	1.1	.2	.1	.3	.6	8.4
				GNP in Constant Dollars (RGNP)[f]					
6	1959–67	(9)	3, 7	1.2	.3	−.7	.4	.7	4.3
7	1962–76	(15)	6, 7	1.2	.3	.4	.4	.5	4.1
8	1969–76	(8)	5, 6, 7, 8	1.2	.7	.7	.3	.3	3.6
9	1969–89	(21)	5, 6	1.1	.1	−.2	.2	.4	3.2
				GNP Implicit Price Deflator (IPD)[f]					
10	1959–67	(9)	3, 7	.6	.1	.1	.2	2.0	1.9
11	1962–76	(15)	6, 7	1.0	0	−.5	0	.8	4.2

12	1969–76	(8)	5, 6, 7, 8	1.4	.1	−.8	.3	.7	5.9
13	1969–89	(21)	5, 6	1.1	.2	−.1	.2	.8	5.6

[a]1: Livingston survey, mean; 2: Mean of eight private forecasts (Harris Bank, National Securities and Research Corp., Conference Board Economic Forum, University of Missouri School of Business, UCLA Business Forecasting Project, *Fortune* magazine, IBM Economic Research Dept., Prudential Insurance Co.); 3: Mean of five forecasts (the first five listed under forecast set 2 above); 4: New York Forecasters Club, mean; 5: ASA/NBER Economic Outlook Survey, median; 6: Council of Economic Advisers (CEA); 7: Research Seminar in Quantitative Economics of the University of Michigan (RSQE Michigan); 8: Wharton School Economic Forecasting Unit, University of Pennsylvania (Wharton).

[b]MAE = $1/n\ \Sigma|E_t|$, where $E_t = P_t - A_t$; P_t = predicted value, A_t = actual value (first estimate).

[c]ME = $1/n\ \Sigma\ E_t$.

[d]Ratio of the mean MAE of forecast (column 3) to the MAE of the corresponding naive model N4 (lines 1–9) or N2 (lines 10–13). N4 projects the moving average of the last four observed changes ($1/n\ \Sigma\ A_{t-i}$, $i = 1, \ldots, 4$). N2 projects the last observed change (A_{t-1}).

[e]Computed from preliminary data (first estimates for year t published in year $t + 1$).

[f]All measures refer to annual percentage changes and are in percentage points.

SOURCES: Zarnowitz (1979, Tables 1–3 for 1953–76; author's files and calculations for 1969–89). See also Moore (1983, Tables 26.3 and 26.4), *Economic Report of the President*, and *Budget of the U.S. Government*.

year for the year ahead. The data come from my earlier and recent work, and their availability dictates the division by subperiods shown in the table. The relative dispersion of the measures across the sets of forecasts and across time tends to be moderate, even when represented by ranges between extreme values.

Thus, the MAEs, in percentage points, average 1.2, 1.2, and 1.0 for GNP, RGNP, and IPD, respectively (column 3). The corresponding figures for the ranges are .4, .4, and .1 (column 4). Reading down the table, a comparison of the successive (mostly overlapping) subperiods suggests that the MAEs may have decreased somewhat for GNP, increased for IPD, and remained remarkably stable for RGNP. But the errors of inflation forecasts increased on average over time much less than the actual inflation rates did, so the accuracy of these forecasts improved greatly in relative terms (compare columns 3 and 8). Such comparisons also suggest a definite reduction in the relative errors for GNP, but a small increase in those for RGNP (note that the average real growth rates decreased slowly between 1959–67 and 1969–89).

It is also instructive to look at ratios of the MAEs of the forecasts to the MAEs of the corresponding extrapolations from selected naive models (column 7).[7] All but one of these relative error measures fall in the range of .3 to .8, indicating the superiority of the forecasts. The single exception is the ratio of 2.0 for the IPD forecasts in 1959–67, a period when projections of last year's rate of change in the price level were surprisingly effective because inflation was unusually low and stable.

In the 1950s and 1960s forecasters generally underpredicted the nominal and real GNP growth rates in years of cyclical expansion (i.e., most of the time). Defined as differences (predicted minus actual values), the errors in these forecasts, therefore, were on the average negative (see the mean errors [MEs] in column 5). The early postwar period enjoyed more real growth than had been expected on the basis of historical experience. Gradually and somewhat belatedly, forecasters learned to be more optimistic. Real GNP increases were strongly underestimated in 1959–67 but overestimated in 1962–76, particularly in 1969–76.

Meanwhile, the IPD forecasts had little if any bias in the period of relative price stability (1959–67); but when inflation was rising and high, it was clearly underpredicted, as in 1962–76 and especially

[7]For GNP and RGNP, projections of four-year trailing moving averages proved relatively effective; for IPD, projections of last year's observed values. Percentage changes based on preliminary data are used.

in 1969–76. Finally, inflation peaked in 1980–81 and decreased markedly in the following five years. Predicted rates moved down with a lag, thus tending to overestimate actual rates. Later inflation increased again but slowly, which was well anticipated.

Real GNP growth was underestimated in 1980, when the recession turned out milder than expected, and in the years of strong recovery and expansion (1983–84, 1988); it was greatly overestimated in 1982, after a severe downturn cut short an unusually weak and brief rise in activity. Thus, as was often observed in the past, the largest errors were associated with business cycle turning points and recessions as well as sharp accelerations and decelerations in inflation (more about this later). However, more than half the time, the annual forecasts for all three variables had only moderate errors of less than one percentage point; also, on the whole the underestimates and overestimates balanced each other well, as can be seen from the results for 1969–89 (lines 5, 9, and 13).

Over some shorter subperiods, the overall mean errors are much larger and so are some of the associated dispersion measures (columns 5–6). Thus it may seem that some of the included forecast sets show undesirable bias (i.e., persistent underprediction or overprediction suggestive of a failure to learn from past errors). The requirements that forecasts should be unbiased and also efficient (uncorrelated with their own errors) are often treated as obvious and minimal in the literature. Yet such requirements are based on assumptions that are only too often falsified in practice, namely that the patterns and relations of the variables concerned are approximately time-invariant and known.

In reality, economic processes are not necessarily stable as they reflect changes in economic institutions, structure, policies, and behavioral rules. Forecasting models and techniques are adapted and altered frequently, and many of the available samples of consistent predictions are too small to establish the existence and to evaluate the importance of any systematic errors. Also, measurement errors may distort and fragment both the time-series data and the related forecasts. For any or all of these reasons, ex post tests can and do find evidences of bias, even in some state-of-the-art predictions from respected sources that ex ante had much professional approval. It seems unlikely that these forecasts were, in fact, systematically deficient in the sense of having persistent yet avoidable errors. I suspect their errors are mainly period-specific and are of the kind that could not be readily detected and corrected at the time the forecasts were made.

Finally, we should note that the forecasts represented in Table 1, though not selected retrospectively for being particularly accurate, have not been randomly chosen either and are certainly superior to many other forecasts. One reason is that I wished to include the important sources with long records such as CEA, Michigan, and Wharton. Another is that the group averages from surveys of forecasters, such as ASA/NBER, are always more accurate over time than most of their individual components (Zarnowitz 1984, 1985). Such averages also conceal the dispersion of errors, which is often large, in the participants' forecasts. The measures presented here and in some other studies suggest that the forecasts by government agencies and teams of econometricians equipped with large-scale models are about as accurate as the survey group averages that represent predominantly judgmental predictions by private business economists.[8]

Tougher Forecasting Tasks and Criteria

In the second post–World War II decade and thereafter, macroeconomic forecasts in the United States grew not only much more abundant but also much bolder. The range of the predicted variables increased greatly as more complete and detailed models were built; it came to include important but very volatile—and hence difficult-to-predict—time series such as corporate profits, housing starts, and inventory investment. Moreover, forecasters were increasingly called upon and able to satisfy the demand for multiperiod predictions of the economy's course. Such forecasts regularly extend over sequences of one to two years and four to eight quarters ahead (some are even longer). This "dynamic" type of forecasting is particularly ambitious. Of course, the computer revolution had much to do with these developments, but so did the advances in economics and statistics, along with the trends in government and business managements.

The average accuracy of forecasts typically decreases as their horizon increases (e.g., GNP is predicted better one quarter than two quarters ahead, better two than three quarters ahead). That is, the MAEs (and RMSEs) rise with each extension of the predictive span, from the current quarter t (for series where preliminary data appear only in the next quarter $t + 1$) through several quarters of the near future. However, the margins by which the absolute or squared errors

[8]Econometric service bureaus usually adjust many predictions generated by their models in attempts to use judgment and up-to-date outside information to correct for errors that an unaided model would commit. The net effects of these constant-term adjustments have been mostly to improve the accuracy of forecasts by compensating partially for the errors in the models and the projected values of the exogenous variables (Evans, Haitovsky, and Treyz 1972; Zarnowitz 1972; McNees 1990).

accumulate over time tend to decrease, and beyond a certain point (often $t + 4$ or $t + 5$) the errors often flatten or vary irregularly around some high plateau.[9] Current information and knowledge may help us forecast $(t + 2)$ better than $(t + 4)$, but we may be about equally ignorant about $(t + 8)$ and $(t + 10)$, for example. These observed tendencies apply to short-term forecasts for a variety of time series; they apply to levels and cumulative absolute and relative changes alike.

Properly understood, the rule "longer forecasts, larger errors" applies to optimal forecasts. It should be a strong regularity—and is. Each of several potential ingredients of a forecast—extrapolation of time series and their relations, external information, or judgment—is subject to a deterioration with the lengthening of the predictive span. Annual forecasts can be viewed as having average spans of about two and a half quarters from the date of issue to the midpoint of their target period; they tend to be about as accurate as comparable forecasts for two or three quarters ahead, less accurate than shorter and more accurate than longer forecasts. Errors in predicting the intrayear quarterly changes often offset each other, which helps considerably the annual forecasts of nominal and real GNP and some of their components. (On the other hand, the multiperiod forecasts of inflation have more cumulation and fewer offsets).

It is not possible or necessary to go beyond this summary of the main features of the relation between the accuracy and the horizon of forecasts. Substantial research has been done on these matters, and its findings are generally consistent with and supportive of the above story (see, e.g., Zarnowitz 1979 and McNees 1988). Clearly, it is much more difficult to predict sequences of quarterly values than single annual values, and the usefulness of point forecasts with long spans and large errors is in doubt. But it is also clear that the tough task of quarter-by-quarter forecasting of the near-term course of the economy in some detail has now become something of a professional routine, presumably in response to the rising demand for just such forecasts. The practice of monthly updating of the forecasts spread concurrently and fast, too: It gives users much fresh information about the outlook for the economy and gives forecasters the opportunity to revise their predictions frequently.

[9] In other words, the errors increase less than in proportion to the horizon (e.g., semi-annual predictions are less than twice as accurate as the annual ones). Indeed, the errors frequently *decrease* with the lengthening horizon for forecasts of growth in the nominal and real aggregates when these are expressed throughout at *annual rates*.

One question that arises at this point is whether the multiperiod predictions are superior to matching mechanical extrapolations from naive and time-series models on the basis of past data available at the time of forecast. Table 2 compares the RMSEs of the forecasts (P) with their counterparts for several naive models (N): projections of last-known levels, changes, and historical averages (N1, N2, and N2*, respectively) and autoregressive extrapolations (N3). The listed ratios RMSE(P)/RMSE(N) are all less than 1, which indicates that the average accuracy of the forecasts is generally higher than that of the naive models. However, in some instances the extrapolations are not much worse than the forecasts: Notably, of the six pre-1970 sets, two have ratios of .9 and higher for the two-quarter span and four for the four-quarter span (lines 1–6, columns 6 and 7). But the best forecasts scored well against the naive models in the early 1970s, a period of great turbulence (lines 7–9). Overall, this fragmentary record illustrates a fairly good but not particularly impressive forecasting performance.

Recently, macro-forecasters have been challenged to exceed higher standards represented by predictions based on vector autoregressive (VAR) models (Sims 1980). In a VAR, each of the variables is predicted by regression on its own lagged values and those of the others; none are exogenous. The number of variables is small, since each is used with several lags. The only use of economic theory and judgment is in choosing the variables. The forecasting process itself is mechanical and replicable. In contrast, econometric forecasting involves exogenous variables that are projected outside the model and, typically, judgmental adjustments to the model outputs of endogenous variables. At least potentially, the roles of both theory and judgment in macroeconometric modeling and forecasting are very large.

Table 3 shows ratios of the average RMSEs of Chase, DRI, and Wharton econometric forecasts to the corresponding measures for predictions from a VAR model by Lupoletti and Webb (1986). The model consists of five variables—RGNP, IPD, the monetary base, the manufacturing capacity utilization rate, and the 90-day treasury bill rate (TBR)—each taken with six lags. It was estimated for 1952:2 to 1969:4, and the obtained coefficients and predictions were then used to forecast each variable for 1970:1 to 1972:2. This procedure was repeated starting with each successive quarter to produce forecasts with horizons of one to six quarters for 1970:1 to 1983:4. Thus, the results are postsample predictions intended to be comparable in this respect to the authentic ex ante forecasts.[10]

[10]However, the data used in the VAR computation were the latest revised estimates

For GNP and RGNP, the RMSE ratios have a range of .73 to .97; for IPD and TBR, half of the ratios favor the forecasts and half favor VAR, but all are close to 1 (see lines 1, 4, 7, and 10 in Table 3). Thus, in most cases the forecasts appear to be more accurate than the much simpler and less expensive VAR projections, but by small margins. The mixed results of comparisons by subperiods (not shown) confirm this conclusion. Perhaps surprisingly, in most cases the relative performance of VAR improved at longer horizons, although prominent econometricians would expect the opposite (Klein 1984, p. 7; Adams 1986, p. 156).

Other evaluations showed that the published forecasts with macroeconometric models more often outperformed univariate ARIMA and multivariate VAR times-series models (McNees 1982; Wallis et al. 1987 for the United Kingdom). In any event, such extrapolations make good competitive standards against which to assess the accuracy of sets of predictions produced by serious and aspiring forecasters.

Overfitting is a major problem for a VAR model that typically includes many terms (their number equals the product, variables times lags) and hence requires estimation of many parameters from a limited amount of data. To avoid or at least reduce this difficulty, Bayesian vector autoregressions (BVAR) use selected restrictions (e.g., that the prior means are one for the coefficients on the first own lag, zero elsewhere, and that the standard deviation of the independent normal distribution for the j-th lag is inversely proportional to j (Litterman 1984, 1986). Thus, the priors contain elements of random-walk models, but the approach is flexible in that it uses alternative proportionality (tightness) specifications and time-varying parameters. McNees (1986) presents detailed comparisons in terms of RMSEs between the regular ex ante BVAR forecasts issued by Litterman in 1980–85 and some of the best-known forecasts by econometricians armed with large-scale models and averages from surveys by business economists. He reports that BVAR forecasts were the most accurate or among the most accurate for RGNP, for the unemployment rate, and for real nonresidential investment and were the least accurate for IPD and (by very small margins) TBR; for GNP, their record was relatively weak over the short spans, strong over the long spans. Table 3 presents a summary of some of this evidence (see lines 3, 6, 9, and 12).

available to the authors, whereas the econometric services used the preliminary estimates available at the time of the forecast. This use could well bias the comparisons considerably in favor of VAR, although Lupoletti and Webb (1986, Table 1 and text, pp. 367–69) present some evidence that this may not be the case.

TABLE 2

SEMIANNUAL AND QUARTERLY MULTIPERIOD FORECASTS OF SIX VARIABLES: COMPARISONS WITH SELECTED NAIVE MODELS OVER SPANS OF ONE TO EIGHT QUARTERS, 1947–75

Line	Variable (Level or Change)[a] (1)		Period Covered[b] (2)	Forecast Sets Covered[c] (3)	Benchmark Model[d] (4)	Span of Forecast, in Quarters: One (5)	Two (6)	Four (7)	Six (8)	Eight (9)
						Ratio: RMSE of Forecast to RMSE of Naive Model[e]				
1	IP	(L)	1947–63	NYFC	N1	—	.71	.80	—	—
2	IP	(L)	1956–63	Fortune	N2*	.68	.62	.92	.80	—
3	GNP	(Δ)	1953–63	Fortune	N3	.70	.74	.96	—	—
4	GNP	(%Δ)	1958–69	IBM	N2	.66	.91	.70	—	—
5	PCE	(%Δ)	1958–69	IBM	N2	.82	.70	.93	—	—
6	GPDI	(%Δ)	1958–69	IBM	N2	.78	.90	.97	—	—
7	GNP	(%Δ)	1970–75	Six Sets	N1	.26	.28	.26	.21	.17
8	RGNP	(%Δ)	1970–75	Six Sets	N1	.45	.55	.67	.69	.73
9	IPD	(%Δ)	1970–75	Six Sets	N1	.32	.34	.40	.43	.48

[a]IP = index of industrial production; PCE = personal consumption expenditures; GPDI = gross private domestic investment; GNP, RGNP, IPD—as used in Table 1; L = level; Δ = change; %Δ = percentage change.
[b]Number of observations per span: line 1, 33; line 2, 13; line 3, 20; lines 4–6, 37, 34, 21 (in columns 5–9, respectively); lines 7–9, 22, 21, 19, 17, and 12 (in columns 5–9, respectively).
[c]NYFC = New York Forecasters Club; Fortune = "Business Roundup" in *Fortune* magazine; IBM = IBM Corp., Economic Research Dept. Six sets = ASA/NBER, BEA, Chase, DRI, GE, and Wharton, mean ratio (BEA = Bureau of Economic Analysis in the U.S. Department of Commerce; Chase = Chase Econometric Associates, Inc.; DRI = Data Resources, Inc.; GE = General Electric Co.; Wharton = Wharton Econometric Forecasting Associates, Inc.). All forecasts are quarterly, except NYFC, which is semiannual.

[d]N1 projects the last known value of the given variable, A_{t-1}; N2, the last known change, ΔA_{t-1}; N2*, the last known average historical change, ΔA. N3 refers to a five-term autoregressive extrapolation (where A_t is regressed on A_{t-i}, $i = 1, \ldots, 5$). Entries in lines 7–9 were computed as Theil's inequality coefficients U ($U = 1$ for the naive model projecting the last known value of the series, here in %Δ).
[e]RMSE = root mean square error. The naive model applied in each line is identified in column 4 (see note d). The entries in columns 5–7 are means of the RMSE ratios for six forecasts each; those in columns 8 and 9 are means of the RMSE ratios for four forecasts each (ASA/NBER and BEA forecasts do not range beyond four future quarters).

SOURCES: Lines 1 and 2, Zarnowitz (1967, Table 18, pp. 100–101); line 3, Mincer and Zarnowitz (1969, Table 1–8, p. 42); lines 4–6, Zarnowitz (1974, Table 6, p. 581); lines 7–9, Zarnowitz (1979, Tables 5, 6, and 7, pp. 18, 20, and 22).

TABLE 3

QUARTERLY MULTIPERIOD FORECASTS OF FOUR AGGREGATE VARIABLES: COMPARISONS WITH VAR AND BVAR MODEL FORECASTS, 1970–85

Line	Period Covered[a] (1)	Forecast[b] (2)	Benchmark Model[c] (3)	Span of Forecast in Quarters: One (4)	Two (5)	Four (6)	Six (7)	Eight (8)
				Ratio: RMSE of Forecast to RMSE of Benchmark Model				
Gross National Product (GNP)								
1	1970–83	Three sets	VAR	.82	.81	.97	.94	—
2	1970–83	Same, TP	VAR	.88	.98	1.48	1.41	—
3	1980–85	Eleven sets	BVAR	.86	.88	.88	.95	1.06
GNP in Constant Dollars (RGNP)								
4	1970–83	Three sets	VAR	.77	.73	.86	.92	—
5	1970–83	Same, TP	VAR	.81	.86	1.04	.96	—
6	1980–85	Eleven sets	BVAR	.93	1.12	1.29	1.74	1.34
GNP Implicit Price Deflator (IPD)								
7	1970–83	Three sets	VAR	.83	.95	1.00	1.00	—
8	1970–83	Same, TP	VAR	.73	.78	.88	.92	—
9	1980–85	Eleven sets	BVAR	.78	.56	.47	.47	.57

90–Day Treasury Bill Rate (TBR)								
10	1970–83	Three sets	VAR	1.11	1.02	.98	.96	—
11	1970–83	Same, TP	VAR	1.19	1.12	1.09	1.10	—
12	1980–85	Eleven sets	BVAR	.90	.91	.97	.83	.77

[a]Numbers of quarterly observations for 1970–83: 53, 52, 50, and 48 in columns 4, 5, 6, and 7, respectively; for 1980–85: 20, 19, 17, 15, and 13.

[b]The three sets include Chase Econometrics, Data Resources, and Wharton Econometric Forecasting Associates (lines 1, 2, 4, 5, 7, 8, 10, and 11). TP refers to forecasts made within two quarters of a business cycle turning point as identified by the National Bureau of Economic Research (NBER). The 11 sets include ASA/NBER median; BEA; Chase; DRI; RSQE, University of Michigan; UCLA (University of California at Los Angeles, School of Business); Wharton; Economic Forecasting Project, Georgia State University; Kent Econometric and Development Institute; Manufacturers Hanover Trust; and Townsend-Greenspan & Co., Inc.

[c]VAR = a five-variable, six-lag quarterly vector autoregressive model. The five variables are the monetary base, the real GNP, the implicit price deflator (all expressed as percentage changes), the manufacturing capacity rate (level), and the 90–day treasury bill rate (level). VAR forecast for 1990:1 is based on data for 1952:2–1969:4, etc. BVAR = Bayesian VAR (see Litterman 1984, 1986).

SOURCES: Comparisons with VAR are in Lupoletti and Webb (1986, Table 3–6, pp. 272–73); those with BVAR are in McNees (1986, Tables 1, 2, 5, and 6, pp. 25–29).

Econometric and judgmental forecasts are commonly believed to have an advantage over time-series forecasts in that they are at least potentially much better equipped to predict turning points. This belief is generally correct for simple forms of extrapolation, but apparently not always for sophisticated time-series models. Table 3 suggests that the BVAR model was often more accurate than the econometricians within two quarters of business cycle turns in 1970–83 (lines 2, 5, 8, and 11).

Overall, the main lesson here is again that the BVAR models include information of predictive value that is not present in the econometric and judgmental forecasts; hence none of these types of predictions are systematically more or less accurate than the others. However, combining the information they contain can improve the forecast. On this point there is general agreement among studies that otherwise offer somewhat different assessments of the relative accuracy of time-series versus econometric forecasts of macroeconomic variables.[11]

Forecasting and the Business Cycle

Forecasters tend to undervalue cyclical movements. The levels of GNP and industrial production, for example, are underestimated most early in a business recovery when growth is particularly strong, less so later when the expansion slows. In general retardations and contractions, the predicted levels as a rule exceed the actual ones, either because the downturn is missed or because the decline turns out to be larger than was forecast.

There are two types of directional errors: a "missed turn," when a turning point in a series occurred but was not predicted, and a "false signal," when a turning point was predicted but did not occur. Since GNP grows most of the time and is expected to, false signals are rare in annual forecasts for this series, but missed turns (as a rule, peaks) are more frequent (Table 4, lines 1–4, column 1). For industrial production (a much more cyclical and volatile series), the percentage of missed turns was smaller and that of false signals much larger; for the GNP implicit price deflator, both relative frequencies were higher yet (lines 1–4, columns 2 and 3).

Quarterly series include many more turning points (both cyclical and minor) than the corresponding annual series. The early multiperiod forecasts of GNP missed most of the declines in the current

[11]Compare McNees (1982, 1986); Lupoletti and Webb (1986); Wallis (1989); Holden and Broomhead (1990, for the United Kingdom); and particularly Fair and Shiller (1990).

TABLE 4

FREQUENCIES OF TURNING POINT ERRORS AND THE ACCURACY OF FORECASTS IN BUSINESS CYCLE EXPANSIONS AND CONTRACTIONS: SELECTED MEASURES FOR FOUR SUBPERIODS, 1947–89

		Annual Forecasts, 1947–65[b]			Quarterly Forecasts of GNP, 1955–63[c]				
Line	Statistic[a]	GNP (1)	IP (2)	IPD (3)	0–1 (4)	1–2 (5)	2–3 (6)	3–4 (7)	5–6 (8)
1	No. of forecast sets	12	11	5	3	3	3	3	3
2	No. of observations	126	127	78	47	45	44	38	39
3	% of actual TP missed	26	15	39	75	83	100	100	100
4	% of predicted TP false	6	18	22	17	13	6	5	—
		Annual Forecasts, 1969–89[d]			Quarterly Forecasts, 1971:2–1985:1[e]				
		GNP	RGNP	IPD	GNP	RGNP	IPD	UR	
5	MAE, b.c. expansions	0.9	0.8	0.9	1.8	1.2	1.1	0.6	
6	MAE, b.c. contractions	1.5	1.9	1.8	3.0	3.0	2.5	1.0	

TABLE 4 (cont.)

FREQUENCIES OF TURNING POINT ERRORS AND THE ACCURACY OF FORECASTS IN BUSINESS CYCLE EXPANSIONS AND CONTRACTIONS: SELECTED MEASURES FOR FOUR SUBPERIODS, 1947–89

		Annual Forecasts, 1969–89[d]			Quarterly Forecasts, 1971:2–1985:1[e]			
Line	Statistic[a]	GNP	RGNP	IPD	GNP	RGNP	IPD	UR
7	ME, b.c. expansions	−0.2	−0.2	0.0	−0.8	−0.4	−0.3	0.1
8	ME, b.c. contractions	1.0	1.4	0.4	−2.4	2.9	−0.8	−1.0

[a]TP = turning points; MAE = mean absolute error; ME = mean error; B.C. = business cycle.

[b]The 12 sets for GNP include 10 forecasts covered in Table 1, line 2 (as listed there in note a) plus averages from two additional large groups, the F. W. Dodge survey of economists and forecasts tabulated annually by the Federal Reserve Bank of Richmond. The 11 sets for industrial production (IP) include the same sources as for the GNP forecasts except Prudential Insurance Co. The 5 sets for IPD include the NICB (Conference Board) Forum, Harris Bank, University of Missouri Business School, UCLA Business Forecasting Project, and National Securities and Research Corp.

[c]Three sets of forecasts are covered: IBM Corp., Economic Research Dept.; New York Forecasters Club; and *Fortune* magazine. 0–1 refers to the change from the current to the next quarter (t to $t + 1$); 1–2 refers to the change from the first to the second future quarter ($t + 1$ to $t + 2$); etc.

[d]Two sets of forecasts are covered: Council of Economic Advisers (CEA) and the ASA/NBER *Economic Outlook Survey*, median (fourth-quarter forecasts for the next year). The years of business contraction, including troughs, are 1970, 1974, 1975, 1980, and 1982 (the other 16 years between 1969 and 1989 are years of business expansion, including peaks).

[e]Five sets of forecasts are covered: ASA/NBER survey median, Chase, DRI, Wharton, and BEA. Expansions, including peaks, cover quarters 1971:2–1973:4, 1975:2–1980:1, 1980:4–1981:3, and 1983:1–1985:1. Contractions, including troughs, cover quarters 1974:1–1975:1, 1980:2–1980:3, and 1981:4–1982:4. UR = unemployment rate.

SOURCES: Lines 1–4, Zarnowitz (1974, Table 7, pp. 588–89; lines 5–8, columns 1–3, author's files and calculations; lines 5–8, columns 4–7, Zarnowitz (1986, Table 2, p. 24).

and next quarters and all of the declines in the more distant quarters ahead (i.e., they falsely predicted rises instead). In contrast, false signals (defined here as predictions of decreases when increases actually occurred) were relatively few and fading with the distance to the target quarter (lines 1–4, columns 4–8).

Real GNP turned down in 1954, 1958, 1970, and 1974. Of the 10 forecasts for these years that were available to me for study, 8 specified continued rises and only 2 succeeded in signaling declines (Zarnowitz 1979, p. 10). Even though they are usually few and far between, cyclical turning-point errors matter greatly because they tend to be exceedingly large. Thus, on average, they are about three times larger than the other errors in forecasts of annual percentage changes in real GNP, as shown in Table 5.

During the 1970s and 1980s, the largest errors in real GNP forecasts occurred in years of business cycle recession and troughs, namely 1970, 1974, and 1982; all were positive. Thus, the forecasts continued to suffer from the failure to predict downturns in aggregate economic activity, even though their relative accuracy improved and the frequency of directional errors decreased compared with the earlier post–World War II period.

A widely observed and strong property of forecasts is that they are more accurate and less biased during periods of business expansion (including peaks) than during periods of contraction (including troughs). When the economy keeps rising, its course is predicted with substantially greater accuracy than when it falls. This result is shown by both the mean absolute errors and the mean errors (disregarding sign), and for both the annual and the quarterly forecasts (Table 4, lines 5–8).

A similar, though somewhat muted, contrast exists between forecasts for the above-average growth phases and those for the below-average growth phases: The former are on average more accurate and less biased than the latter.[12] In sum, large errors tend to cluster around and immediately after business cycle turns and growth cycle turns, especially peaks (growth cycles are major fluctuations in trend-adjusted aggregates).

Note that the meaning of these results is not simply that the forecasting failures are due to large unanticipated shocks, for the latter can and often do occur under any economic conditions. The concentration of large errors during slowdowns and contractions cannot be explained away by a general reference to random outside disturbances. The economy is particularly vulnerable in these business

[12]For evidence, see Zarnowitz (1986, Table 2).

TABLE 5

FORECAST ERRORS OF ANNUAL PERCENTAGE CHANGES IN REAL GNP

	Underestimates		Overestimates		Turning Point Errors	
	No.	MAE	No.	MAE	No.	MAE
Five sets, 1959–76[a]	33	1.1	21	.9	8	2.8
CEA, 1969–89	8	.9	9	.8	3	2.3
ASA/NBER, 1969–89	9	1.2	8	.5	3	2.8

[a]Includes the first 5 sets listed under code 2 in Table 1, note a. For further details, see Zarnowitz (1979, Table 2, col. 1).

cycle phases owing to a gradual accumulation of various stresses and imbalances, and it is very difficult to predict just when these phenomena will culminate. Also, few forecasters take the risk of signaling a recession prematurely ahead of others; the costs of such prediction to themselves and their customers can be quite high. In contrast, forecasts of a recovery are always welcome and often accepted on the basis of early signs of improvement. The peak errors show up during the recession and slowdown periods, the generally smaller trough errors show up during the recovery and speedup periods.

The most influential private forecasts are now issued monthly for sequences of several quarters ahead; government forecasts no doubt are adjusted just as frequently, though only for internal uses, not public knowledge. How early can the alert producers of such predictions foresee or detect major events such as the turning points in business cycles, growth cycles, and inflation fluctuations? Experience varies but the probable lead times at peaks, if any, are short. For example, the first forecasts of a downturn in 1973 were coincidental with the onset of the recession in the fall; the many predictions of a peak in 1979 found much support in preliminary data but were not confirmed until 1980:1; the mid-1981 peak was widely missed; and few predicted a decline in 1990 before August, though soon thereafter most forecasters agreed that a recession was under way.[13]

Most of the recent business contractions were preceded by fairly long slowdowns in aggregate economic activity (1957, 1969, 1973, 1979, 1989–90). A number of leading indicators and corresponding composite indexes declined or flattened early on each of these occasions, providing early signals that the economy was weakening. Many forecasters, monitoring these developments, promptly recognized the slowdowns but discounted the associated recession risks. As a result, the forecasts, like the indicators, tend to have a better record of timely prediction for slowdowns (growth cycle peaks) than recessions (business cycle peaks).

At business cycle troughs, the leads of the indicators are generally much shorter than at peaks but are also much less variable. Thus, forecasters may trust these signals more, but they are too often ready to predict the end of a recognized recession much earlier. So, the prevailing view in spring 1974 (after the oil embargo ended) was that the recession was about over, but that view was premature by almost a year. Similarly, most forecasters expected the 1981–82

[13]For detail, see McNees (1990, pp. 159–67) and (on the 1990 peak forecasts) Zarnowitz (1990, 1991).

contraction to last two to three quarters rather than the actual five quarters. (But this error was in part related to the opposite error of overestimating the length of the 1980 recession, which was unusual in lasting only six months.)

The turning points marking the major rises and declines in inflation have been for the most part poorly predicted, with forecasts lagging behind the actual values much as simple extrapolations would. A detailed analysis of a large number of forecasts for several variables found inflation errors particularly troublesome; so have some other studies (Zarnowitz 1985; Holden and Peel 1985 for the United Kingdom).

Conclusion

There is much disenchantment with macroeconomic forecasting. The difficult question is "How much is due to unacceptably poor performance and how much to unrealistically high expectations?" I would argue that the latter is a major factor. Economists were held in high repute during the 1960s, probably in large part because the macro-forecasts looked good then, and high growth and prosperity prevailed for some time with inflation still well restrained. But it is relatively easy to achieve a respectable forecasting record in times of continuing expansion. Later, when inflation accelerated, when serious recessions reappeared, and when long-term growth of productivity and total output slackened, the errors of macro-models and macro-forecasts received increased public attention—as did the old and new controversies among the economists. The reputation of the profession suffered and, perhaps worse yet, the interest of academic economists in forecasting, never very strong, weakened still more. Yet the performance of professional economic forecasters, when assessed in proper relative terms, has been considerably better in recent times than in the earlier post–World War II period. What happened is that the improvements fell short of enabling the forecasters to cope with the new problems they faced.

As a practical activity, the results of which are marketed, recorded, researched, and tested, macroeconomic forecasting is very young by any standard. There is little doubt that it will always disappoint the hopes of many, but also a high probability that it can be developed well beyond its present early stage. If macroeconomics has a long way to go, as I believe to be the case, then macro-forecasts, too, should still be far from the limits on their improvability, even if such limits were to prove much narrower than early enthusiasts thought.

Progress in forecasting will require chiefly better data and models, but also improvements in time-series analysis, econometric methods, cyclical indicators, and anticipations surveys. These essentially complementary tools should be used efficiently as such, not as competitors or substitutes. We can expect the advances to prove difficult and slow with setbacks along the way. Indeed, some large banks and industrial companies have sharply reduced or even liquidated their economic staffs in recent years. But this move can hardly mean that these organizations have suddenly discovered that they can do without forecasts of important aspects of aggregate economic activity on which their own business prospects may depend critically. More likely, they decided that other ways to acquire such forecasts (e.g., subscription to outside services or surveys) are more economical than in-house production. The predictive needs of decisionmakers who are necessarily future-oriented are not reduced by the perceived shortcomings of past forecasts. This view applies to government policymakers as well as to private agents.

However, there are ways to reduce one's dependence on forecasts to a degree, and the incentives to use them are presumably greater the more fallible the forecasts appear. One partial substitute for the lacking foresight is readiness to respond promptly and flexibly to unforeseen changes. For example, private reactions to economic fluctuations may involve employment smoothing through labor hoarding along with production smoothing through changes in unfilled order backlogs.[14] An example of a government policy that is relevant here would be job creation through countercyclical public works or public service employment programs, which would draw on an advance preparation of a backlog of useful projects to be activated as needed (compare Council of Economic Advisers 1954, p. 123). But fiscal policies of this kind were more often pro-cyclical than counter-cyclical because of tardiness and poor planning (see Zarnowitz and Moore 1982, pp. 57–59).

The principal proposed alternative to discretionary government actions that must rely on forecasts is to consistently follow a stable policy rule that would call either for no response or for a predetermined response to changes in the economy (e.g., a fixed growth rate for some controlled monetary aggregate or a rate varied as a function of, say, the observed inflation). Policy rules, it is often argued, can be expected to have positive stabilizing effects on private expectations

[14]Also, theoretically, through changes in inventories of produced goods; but empirically, inventories do not seem to be used to accommodate cyclical fluctuations in demand (see, e.g., Blinder 1986).

and to discipline the authorities that may otherwise be tempted to engage in shortsighted attempts to overstimulate the economy by inflationary policies. Thus, the deficiency of forecasts is not the only argument used in favor of the rules; nor is it necessarily the main one. Monetary control could be poor even with accurate forecasts because of inconsistent and inefficient procedures with regard to the instruments and targets of the policy.

Are even the best available forecasts inadequate to be a basis for satisfactory policy decisionmaking? This question probably does not have a single, clear-cut answer at this time. There is evidence that the government has no substantial and lasting informational advantage; notably, the CEA and the Federal Reserve forecasts are about as accurate as the state-of-the art private forecasts (as indicated by our results and other studies; see, e.g., Meltzer 1991, pp. 30–32). Certainly, the forecasts cannot support "fine-tuning," that is, keeping the economy always very close to full employment; but this would not be a realistic goal for balanced policies even if macroeconomic forecasting were in far better shape than it presently is. Yet forecasts should be sufficient most of the time to assist in the pursuit of reasonable policy objectives: preventing or at least effectively combating persistent high unemployment and persistent high inflation.

The main defects of macro-forecasts from the point of view of policy are the errors of missing cyclical turns and shifts in the average rates of inflation. Major reductions in such errors should rank high on the agenda of economists.

References

Adams, F. Gerard. *The Business Forecasting Revolution: Nation-Industry-Firm*. New York: Oxford University Press, 1986.

Batchelor, Roy A., and Dua, Pami. "Forecaster Ideology, Forecasting Techniques, and the Accuracy of Economic Forecasts." *International Journal of Forecasting* 6 (1) (1990): 3–10.

Blinder, Alan. "Can the Production Smoothing Model of Inventory Behavior Be Saved?" *Quarterly Journal of Economics* 101 (August 1986): 431–53.

Council of Economic Advisers. *Annual Report*. Washington, D.C.: Government Printing Office, 1954.

Evans, Michael K.; Haitovsky, Yoel; and Treyz, George I.; assisted by Vincent Su. "Analysis of the Forecasting Properties of U.S. Econometric Models." In *Econometric Models of Cyclical Behavior*, vol. 2, pp. 949–1158. Edited by B. G. Hickman. New York: Columbia University Press for the National Bureau of Economic Research (NBER), 1972.

Fair, Ray C., and Shiller, Robert J. "Comparing Information in Forecasts from Econometric Models." *American Economic Review* 80 (3) (June 1990): 375–89.

Holden, K., and Broomhead, A. "An Examination of Vector Autoregressive Forecasts for the U.K. Economy." *International Journal of Forecasting* 6 (1) (1990); 11–23.

Holden, K., and Peel, D. A. "An Evaluation of Quarterly National Institute Forecasts." *Journal of Forecasting* 4 (1985): 227–34.

Klein, Lawrence R. "The Importance of the Forecast." *Journal of Forecasting* 3 (1) (January–March 1984): 1–9.

Litterman, Robert B. "Specifying Vector Autoregressions for Macroeconomic Forecasting." Staff Report 92. Federal Reserve Bank of Minneapolis Research Department, 1984.

Litterman, Robert B. "Forecasting with Bayesian Vector Autoregressions—Five Years of Experience." *Journal of Business and Economic Statistics* 4 (1) (January 1986): 25–38.

Lupoletti, William M., and Webb, Roy H. "Defining and Improving the Accuracy of Macroeconomic Forecasts: Contributions from a VAR Model." *Journal of Business* 59 (2) (April 1986): 263–85.

McNees, Stephen K. "The Forecasting Record for the 1970s." *New England Economic Review* (September/October 1979, corrected version): 1–21.

McNees, Stephen K. "The Recent Record of Thirteen Forecasters." *New England Economic Review* (September/October 1981): 5–21.

McNees, Stephen K. "The Role of Macroeconometric Models in Forecasting and Policy Analysis in the United States." *Journal of Forecasting* 1 (1982): 37–48.

McNees, Stephen K. "Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts." *Journal of Business and Economic Statistics* 4 (1) (January 1986): 5–15.

McNees, Stephen K. "How Accurate Are Macroeconomic Forecasts?" *New England Economic Review* (July/August 1988): 15–36.

McNees, Stephen K. "Man vs. Model? The Role of Judgment in Forecasting." *New England Economic Review* (July/August 1990): 41–52.

Meltzer, Allan H. "The Fed at Seventy-Five." In *Monetary Policy on the 75th Anniversary of the Federal Reserve System*, pp. 3–83. Edited by M. T. Belongia. Proceedings of the 14th Annual Economic Policy Conference of the Federal Reserve Bank of St. Louis. Boston: Kluver, 1991.

Mincer, Jacob, and Zarnowitz, Victor. "The Evaluation of Economic Forecasts." In *Economic Forecasts and Expectations*, chap. 1. Edited by J. Mincer. New York: NBER, 1969.

Moore, Geoffrey H. "The President's Economic Report: A Forecasting Record." In *Business Cycles, Inflation, and Forecasting*, chap. 26. Edited by G. H. Moore. Cambridge, Mass.: Ballinger Publishing Co. for the NBER, 1983.

Sims, Christopher A. "Macroeconomics and Reality." *Econometrica* 48 (1980): 1–48.

Su, Vincent, and Su, Josephine. "An Evaluation of the ASA/NBER Business Outlook Survey Forecasts." *Explorations in Economic Research* 2 (1975): 588–618.

Wallis, K. F. "Macroeconomic Forecasting: A Survey." *Economic Journal* 99 (March, 1989): 28–61.

Wallis, K. F.; Fisher, P. G.; Longbottom, J. A.; Turner, D. S.; and Whitley, J. D. *Models of the U.K. Economy: A Fourth Review by the ESRC Macroeconomic Modeling Bureau.* Oxford: Oxford University Press, 1987.

Zarnowitz, Victor. *An Appraisal of Short-Term Economic Forecasts.* New York: NBER, 1967.

Zarnowitz, Victor. "New Plans and Results of Research in Economic Forecasting." *Fifty-Fifth Annual Report.* New York: NBER, 1971.

Zarnowitz, Victor. "Forecasting Economic Conditions: The Record and the Prospect." In *The Business Cycle Today*, pp. 183–239. Edited by V. Zarnowitz. New York: NBER, 1972.

Zarnowitz, Victor. "How Accurate Have the Forecasts Been?" In *Methods and Techniques of Business Forecasting*, pp. 565–96. Edited by W. F. Butler, R. A. Kavesh, and R. B. Platt. Englewood Cliffs, N.J.: Prentice-Hall, 1974.

Zarnowitz, Victor. "An Analysis of Annual and Multiperiod Quarterly Forecasts of Aggregate Income, Output, and the Price Level." *Journal of Business* 52 (1) (January 1979): 1–33.

Zarnowitz, Victor. "The Accuracy of Individual and Group Forecasts from Business Outlook Surveys." *Journal of Forecasting* 3 (1) (January–March 1984): 11–26.

Zarnowitz, Victor. "Rational Expectations and Macroeconomic Forecasts." *Journal of Business and Economic Statistics* 3 (4) (October 1985): 293–311.

Zarnowitz, Victor. "The Record and Improvability of Economic Forecasting." *Economic Forecasts: A Worldwide Survey* 3 (12) (December 1986): 22–3.

Zarnowitz, Victor. Reports in *Economic Forecasts: A Monthly Worldwide Survey* 7 (6–12) (June–December 1990) and 8 (1) (January 1991).

Zarnowitz, Victor, and Moore, Geoffrey H. "Sequential Signals of Recession and Recovery." *Journal of Business* 55 (January 1982): 57–85.

Criteria for Judging Macroeconomic Forecasting

Allen Sinai

Victor Zarnowitz concentrates on a limited dimension of the question that titles his paper, "Has Macro-Forecasting Failed?" That dimension is macroeconomic forecast accuracy and the forecast performance of macroeconomic projections by different types of forecasters, different forecast methods, and across different periods. As a paper on the forecast accuracy of macroeconomic projections and the comparative forecasting record of various methods of macroforecasting, the Zarnowitz paper has much to offer on four counts: (1) an evaluation of the record from 1969 to 1989; (2) how judgment-type forecasting, model-based forecasts, and time-series methods do vis-a-vis one another; (3) some myths surrounding common perceptions about macro-forecasting; and (4) some of the underlying nitty-gritty of how macro-forecasting is done.

Zarnowitz does perpetuate a few incorrect notions about macroforecasts and macro-forecasting, however. And there is excessive material that is extraneous to the purpose of the paper and to the limited aspect of the question he, in fact, examines. But, by and large, if one takes forecast accuracy and performance as the major criteria of whether macro-forecasting has failed, the paper makes a significant contribution.

Has Macro-Forecasting Failed? How Can We Tell?

The major deficiency is that Zarnowitz never directly comes to grips with the question that is posed: Has macro-forecasting failed?

Cato Journal, Vol. 12, No. 1 (Spring/Summer 1992).

The author is Executive Vice President and Chief Economist of The Boston Company, Inc., President and Chief Executive Officer of The Boston Company Economic Advisors, Inc., and an Adjunct Professor in the Lemberg Program at Brandeis University.

A full and comprehensive list of criteria relating to the methods, uses, and accuracy of macro-forecasting is not laid out, on which the pluses and minuses of macro-forecasting could be examined. There is casual reference to some additional criteria, but that reference disappears as Zarnowitz pursues the accuracy issue and the empirical data in his very impressive database of historical macro-forecasts from different types of forecasters and different forecast organizations.

Even on the accuracy criterion, though, I could not tell what Zarnowitz concluded from his examination of the evidence, Unless, as he hinted, his task was simply to lay out the facts on forecast accuracy and methods as information input for the reader to render a judgment. But my real criticism is that to evaluate the failures and successes of macro-forecasting and the methods various practitioners use to advance the activity, a much broader kind of examination than the one taken by Zarnowitz is required.

The historical record of macro-forecasts and macro-forecasting is a necessary dimension on which to consider whether macro-forecasting has failed. But it is far from sufficient in light of the multitude of uses and applications—both actual and potential—of macro-forecasting. Indeed, it could be argued that in a microeconomic, bottom-line, decisionmaking context, the value of macro-forecasting and its success or failure goes far beyond the accuracy of point forecasts, or the rate of growth of GNP, or the rate of inflation and their turning points.

Point forecasts based on conditional assumptions and conditional probabalistic models must always be wrong. Focusing on accuracy, therefore, tends to distract and obfuscate the best and most profitable uses and decisions that relate to macro-forecasting in government, business, and financial institutions. For all its errors in accuracy and difficulties that are presented in use, would we ask whether meteorology has failed? Macro-forecasting, I would submit, does better and is more useful and necessary for business decisions than is meteorology.

Accuracy and Other Criteria

If I levy this criticism, how might macro-forecasting be judged as a success or failure? Without going into detail, let me list and comment briefly on some of the criteria I think are valid in answering the question, "Has macroforecasting failed?"

The Accuracy Issue

First, there is the accuracy issue, which is important—a question of levels, changes, and turning points in forecasted variables compared with actual values. The postwar record here is mixed. Zarnowitz does very well in presenting it plus helping users of the macro-forecasts understand the biases so that appropriate adjustments can be made. For example, it should be understood and well known that macro-forecasting always understates upswings and downswings. It is the nature of the beast, and Zarnowitz brings that point out well. The data are seasonally adjusted, models tend to add smoothing, and forecasters smooth the smooth results of the models and the data. It is really in the culture of what goes on in macro-forecasting that one gets underestimates of both the upswings and downswings. But forecasts tend to improve over time, as Zarnowitz notes, especially in relation to the magnitudes and difficulties at the time of forecasting. The oil shock of 1973–74 was a new event: it increased both inflation and unemployment. Was macro-forecasting a failure because policymakers did not fathom the implications of the oil shock or engage in the right policy response? One might come to that conclusion if the focus is on forecast accuracy and point forecasts, which distract users from maximizing for decisions the information content of the forecasts.

Wrong forecasts, however, do not necessarily mean bad decisions or bad bottom-line results. Point forecasts and forecasts of turning points will always be wrong, since neither the science of macro-forecasting nor its practitioners can really produce point forecast results on turning points or on specific variables. Once the limitations of macro-forecasting are recognized, the informational use of such forecasting is enhanced.

Other Criteria for Judging Macro-Forecasting

There are many other criteria besides accuracy for judging macro-forcasting. One criterion is the use of macro-forecasts as a backdrop for micro-decisions. It is like forecasting the weather. Macro-baseline expectations and alternative prospects are essential ingredients for virtually all micro-unit decisions and decisionmakers. That is an intelligent and successful use of macro-forecasting. Decisionmakers in government and business use macro-forecasts as benchmarks, but those benchmarks are revised as new information comes in. Again, the point forecasts cannot be taken literally. They can be used as initial guides to macroeconomic policy and to form expectations in financial markets. When a forecast fails, or is not accurate, policymakers can react.

Another criterion by which to judge macro-forecasting is how well it enhances our understanding of the macroeconomy. Today macro-results and model results of various policies are examined thoroughly through macro-forecasting. This can make for less chance of a really bad and poorly timed policy action. Policies are now examined in a way that never was true before. The examination of the stock market crash of October 1987 indicated no recession for most of those who ran the numbers through systematic models. For me, when I did that, that was a surprise, but a result from the discipline of the macro-forecasting process regardless of any forecast.

A further dimension is the increase in current-quarter monitoring as part of macro-forecasting. There is now almost continuous forecasting and monitoring of the near term as an input to decisions, that is, as a monitor in terms of deviations from the forecast or from expectations. That development is more important for bottom-line decisions and good policymaking than, I believe, the forecasts themselves.

Finally, in thinking about relevant criteria for determining whether macro-forecasting has failed, one must consider what can be reasonably expected from the forecasts and the forecasters. Perfect accuracy is impossible and should not be the sole criterion upon which to judge the success of macro-forecasting.

Conclusion

Zarnowitz has made a significant empirical and scientific contribution on the accuracy record of macro-forecasts in the postwar period. He also has increased our understanding of how macro-forecasts can be used. But I do not think he has begun to approach the more general title of his paper. I would say that over the postwar period, allowing for the criteria I have laid out, there are pluses and minuses—and I would be hard put to say that macro-forecasting has failed once all of its uses, actual and potential, are considered.