

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Business Cycles: Theory, History, Indicators, and Forecasting

Volume Author/Editor: Victor Zarnowitz

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-97890-7

Volume URL: <http://www.nber.org/books/zarn92-1>

Conference Date: n/a

Publication Date: January 1992

Chapter Title: The Record and Improvability of Economic Forecasting

Chapter Author: Victor Zarnowitz

Chapter URL: <http://www.nber.org/chapters/c10389>

Chapter pages in book: (p. 519 - 534)

---

# 18 The Record and Improvability of Economic Forecasting

## 18.1 Questions and Problems

The question “Is better forecasting possible?” would seem to be of critical importance to both makers and users of macroeconomic predictions. Indeed, the ability to produce accurate predictions of the course of the economy in the near-term future is probably the main criterion by which the public judges the usefulness of our entire profession. It is true that this popular standard fails to discriminate between wrong specifications of economic models and wrong choices of assumptions about outside events, whereas one may argue that an economist should be held mainly responsible for the former rather than the latter source of forecast errors. But most economists would agree that the proper test of the practical aspects of their expertise consists in how well they can predict or “explain” postsample data.

Critics often assert that the economic forecasts generally are poor. However, it is not clear what standards they apply and whether such complaints represent more than casual opinions. Large errors can occur for a variety of reasons and need not be either systematic or symptomatic of forecasters’ inability.

Logically, the inquiry into the improvability of forecasting should start with some prior questions: How accurate have the forecasts been on the average in the past? What are the sources and characteristics of superior forecasts? For several reasons, however, these seemingly simple questions lack unique and conclusive answers.

1. The forecasts must be explicit, verifiable, and sufficient to permit a responsible appraisal. But the recorded history of macroeconomic forecasting is of recent origin. Time series on specific, quantitative, and comparable predic-

tions are as a rule short. Few forecasters have been active consistently over many years; many offer only small samples of observations with isolated hits or misses that could be largely due to chance.

2. Some periods are easier to forecast than others. For example, once it is clear that a recession has just ended, it is a rather safe bet that the recovery will continue in the months immediately ahead, but just when a mature expansion will end is usually quite difficult to anticipate.

3. Some variables are easier to forecast than others. In general, the trend-dominated and smooth series are better predicted than the cyclical and volatile series. Forecasting models differ greatly in size and complexity—the number and composition of endogenous and exogenous variables. It is difficult to make dependable comparisons across such models.

4. Economic agents generally use the forecasts to help formulate and improve their plans and decisions. They expect that the value of the resulting reductions in their errors will tend to exceed the effective costs to them of producing or acquiring the forecasts. However, these costs and returns are typically difficult to estimate and unknown to an outside analyst. Users have different needs, skills, and preferences (“loss functions”). The *size* of forecasting errors may not be sufficient to determine their *consequences* for the decisions based on the forecasts.

5. Different summary measures of error may lead to different appraisals of a given set of forecasts. The results will depend on whether the averages are based on absolute or squared errors; on whether the errors are computed for predictions of levels or changes; and on the importance of measurement errors and the treatment of data revisions. Absolute accuracy measures, which show deviations from the obviously unattainable state of perfection (zero errors), need to be complemented with relative accuracy measures, which compare forecasts from different sources or of different types. Here the standard is often some objective “benchmark” model, for example, low-cost extrapolations of the own history of the target series. The optimal standards vary with the properties of the time series in question.

6. Ideally, forecasts should be unbiased, that is, have random, nonautocorrelated errors averaging 0, since forecasters should use the available information to eliminate all avoidable systematic errors. But success in this endeavor requires sufficiently large samples of comparable predictions on series generated by sufficiently stable processes. If these conditions are not met, the apparent bias may be spurious. A different reason for the same result may lie in asymmetric loss functions.

7. Macroeconomic forecasts vary greatly with respect to the relative roles of model and judgment, but in practice inevitably include elements of both. There is no way to avoid judgment in the choice of the model itself and on how the modeled regularities of the sample period are to be modified in light of new and external events. Uses of objective, reproducible methods offer valuable opportunities for learning, the results of which can be recorded and

published. This advantage is not provided by those forecasters who do not disclose their assumptions and techniques or models. Experienced judgment may be the most valuable property of a forecaster but it is not something that can be readily transmitted to others.

In sum, the quality of forecasts is a relative and multidimensional concept. Forecasts vary in many ways: by source, techniques, variables and periods covered, timing, and horizon. It is generally difficult to allow for these differences so as to make meaningful comparisons across forecasters and over time, even after the event, with data on the corresponding actual values on hand.

The history of modern forecasting overlaps the “information revolution” of the last 30 years, a period of rapid expansion in the scope and content of economic data, measures, and literature. The process was (and is) the result of a number of interacting developments on both the demand side and the supply side: advances in data collection and processing, in economic theory, statistics, and econometrics; the accelerating power of the computer; the spread of modern management techniques propelled by competition; the growing size and planning requirements of governments. As usual in times of revolutionary change, great expectations were born. Some of these promised too much. This certainly applies to the notion of a road to dependable business and economic policies built by a new science of optimum forecasting.

## 18.2 Some Evidence and Interpretations

Is there a way to address what appears to be a complex question of *trends* in forecasting accuracy without getting bogged down in the many differences among forecasters, techniques, models, variables, horizons, and periods covered? I hope to show that the answer is a qualified yes. The problems discussed above will not be resolved but the complications they pose can be reduced by the design of the study, and some limited but pertinent results can be obtained from the available record.

### 18.2.1 Annual Forecasts: Comparisons across Time and Sources

Table 18.1 arrays, by common coverage in time, measures of average error without regard to sign for a large collection of annual *ex ante* forecasts of nominal and real growth and inflation. These are the longest authenticated time series of this kind that could be collected, but they reach back only to 1953 (for GNP) and 1959 (real GNP and the implicit price deflator).<sup>1</sup> There are gaps and overlaps in this compilation that one would wish away, but they reflect the availability of the data and could not be avoided. The forecasts are roughly comparable in timing, most having been made in October or November of year  $t$  for the year  $t + 1$ , that is, before the publication of the first official GNP estimates for the last quarter of year  $t$ .

1. See notes to table 18.1 for sources of the forecasts covered and references to related studies.

Table 18.1

**Summary Measures of Error for Annual Forecasts of Percentage Changes in  
Aggregate Income, Output, and the Price Level, 1953-1984.**

Line (1)	Period and No. of years Covered (2)	Mean Absolute Error of Forecasts (MAE)							MAE		Relative Error (10):(11)
		LIV <sup>a</sup> (3)	SPF <sup>b</sup> (4)	NYF <sup>c</sup> (5)	ERP <sup>d</sup> (6)	ANB <sup>e</sup> (7)	MIM <sup>f</sup> (8)	WHM <sup>g</sup> (9)	Mean <sup>h</sup> (10)	XP <sup>i</sup> (11)	
<i>Growth Rate of Gross National Product (GNP)</i>											
1	1953-76 (24)	1.6	1.2						1.4	2.3	0.6
2	1956-63 (8)	1.7	1.4	1.7					1.6	1.9	0.8
3	1963-76 (14)	1.1	0.9		0.9		1.3	0.8	1.0	1.8	0.6
4	1969-76 (8)	1.0	0.6		0.8	1.0	1.0	0.9	0.9	2.0	0.5
5	1977-84 (8)				1.8	1.5	2.0	1.6	1.7	2.8	0.6
<i>Growth Rate of GNP in 1972 Dollars (RGNP)</i>											
6	1959-67 (9)		1.3				1.0		1.2	1.7	0.7
7	1962-76 (15)				1.1		1.4		1.2	2.6	0.5
8	1969-76 (8)				1.2	1.0	1.6	0.9	1.2	3.6	0.3
9	1977-84 (8)				1.2	1.0	1.0	1.0	1.0	3.2	0.3
<i>Rate of Inflation in the GNP Implicit Price Deflator (IPD)</i>											
10	1959-67 (9)		0.6				0.7		0.6	0.3	2.0
11	1962-76 (15)				1.0		1.0		1.0	1.3	0.8
12	1969-76 (8)				1.4	1.3	1.4	1.4	1.4	2.0	0.7
13	1977-84 (8)				0.9	1.0	1.2	1.2	1.1	1.3	0.8

<sup>a</sup>Based on surveys conducted by Joseph A. Livingston, syndicated columnist. Published in the *Philadelphia Bulletin* and *American Banker* and, in recent years, in the *Philadelphia Inquirer*. Of the semiannual surveys, only the end-of-year ones are used here; questionnaire typically mailed in November and results published in December. Coverage 44-62 persons.

<sup>b</sup>Mean of end-of-year forecasts from the following sources: (1) *Fortune* magazine ("Business Roundup"); (2) Harris Bank; (3) IBM Economic Research Department; (4) National Securities and Research Corporation; (5) NICB, now Conference Board "Economic Forum"; (6) R. W. Paterson, University of Missouri; (7) Prudential Insurance Company of America; (8) UCLA Business Forecasting Project. The earliest of these predictions were made in October; the latest in January. Most of these forecasts are quarterly. For studies of these data through 1976, see Zarnowitz 1967, 1972b, 1974, and 1979 (see chapter 14, this volume).

<sup>c</sup>Group mean forecasts from the New York Forecasters Club. Of the semiannual forecasts, only the end-of-year ones are included. Coverage: 31-39 individual respondents. Dates: 1956-58, October; 1959-63, December. Collected through 1963 and analyzed in Zarnowitz 1967, 1972b, 1974, and 1979 (chapter 14, this volume).

<sup>d</sup>Annual forecasts by the Council of Economic Advisers (CEA) as stated in the *Economic Report of the President*, published as a rule in January. Often midpoints in the relatively narrow range, in a few cases interpolated and checked with the source for approximate accuracy. See Moore 1969a, 1977d, 1982; Zarnowitz 1972b, 1979 (chapter 14, this volume); Fellner 1976; McNees 1977.

<sup>e</sup>Source: Quarterly releases by the American Statistical Association (ASA) and the National Bureau of Economic Research (NBER), published by ASA in *AmStat News* and by NBER in *Explorations in Economic Research* and, more recently, *NBER Reporter*. Median forecasts from the November surveys only are used. Coverage varied between 25 and 84, but mostly 30-50. See Zarnowitz 1967, 1969b, 1972b, 1974, and 1979 (chapter 14, this volume), 1984a, 1985a; Mincer and Zarnowitz 1969; Moore 1969a; Su and Su 1975; McNees 1973, 1974, 1975, 1976.

<sup>f</sup>Forecasts from the Research Seminar in Quantitative Economics (RSQE) of the University of Michigan. Published quarterly (initially, three times per year). Included here are the forecasts released in connection with the University of Michigan annual "Conference on the Economic Outlook," dated as a rule in November. Based on several working models; see Suits 1962; Hymans and Shapiro 1970, 1974.

<sup>g</sup>Source: *Wharton Economic Newsletter*, Econometric Forecasting Unit, Wharton School of Finance and Commerce, University of Pennsylvania. Published quarterly; the forecasts used here are end-of-year, as a rule dated in November. Based on a series of Wharton models (see Evans and Klein 1967; Evans, Klein, and Saito 1972; McCarthy 1972b; Duggal et al. 1974).

<sup>h</sup>Mean of the entries in cols. 3-9.

Table 18.1 Continued

Extrapolative benchmark forecasts. For GNP and RGNP (lines 1–9) assumes that next year's percentage change will be the same as the average percentage change in the four previous years. For IPD (lines 10–13) assumes that next year's percentage change will be the same as that of previous year. The actual changes are based on the first official estimates following the year for which the forecast was made. See text.

Table 18.1 has a highly diversified coverage. It includes averages from regular surveys of professional forecasters (cols. 3, 5, and 7), various predictions selected for early and consistent coverage (col. 4), forecasts by the successive teams of presidential economic advisers (col. 6), and forecasts by two econometric service bureaus (cols. 8 and 9). Individual and collective judgments, informal and formal approaches, small and large models—all of these are well represented. Each line refers to a period that covers a variety of business conditions.

The mean absolute error (MAE) measures assembled in table 18.1 display no systematic upward or downward trends, as can be seen by comparing the entries within each of columns 3–9 for the individual forecast sets. The overall means in column 10 convey the same message. True, errors in the annual predictions of nominal GNP growth rates were on the average larger in the last eight complete years than in the eight previous years, for example, but the opposite applies to the predictions of real GNP growth and inflation.

To allow for any changes in the means of the predicted series across the periods covered, benchmark MAE measures were computed for selected naive models: 4-year moving-average extrapolations for the annual percentage changes in GNP and RGNP, last-year extrapolations for those in IPD.<sup>2</sup> Comparisons of the forecast errors (cols. 3–10) with the naive extrapolation (XP) errors (col. 11) show that the former are in all but one case (line 10) substantially smaller than the latter. Moreover, the relative errors, that is, ratios of MAE for the forecasts to the corresponding MAE-XP measures (see col. 12), show a tendency to decline between the earlier and the more recent periods. On this criterion, then, one would conclude that the annual forecasts of nominal and real GNP growth may have actually improved, at least since the late 1950s. (For inflation, the evidence is weaker, as shown in lines 10–13.)

It is also interesting to note that, consistent with several studies such as McNees 1975 and 1976 and Zarnowitz 1967, 1972b, and chapter 14, this volume, it is difficult to detect systematic differences in accuracy among the well-known professional forecasters. In general, the MAE statistics for the forecast sets included in table 18.1 do not differ much. This is well illustrated by the following tabulation:

2. These simple models perform relatively well for the respective variables, and more elaborate time-series models are neither needed nor properly applicable here. Annual data comparable to those available to the forecasters are short also; see table 18.1, n. i.

Entries in table 18.1, cols. 3–9, for

	<i>GNP</i>	<i>RGNP</i>	<i>IPD</i>
Mean ( $\bar{x}$ )	1.2	1.1	1.0
Standard deviation ( $s$ )	0.4	0.2	0.3
Fraction within $\pm 1 s$ around $\bar{x}$	15/20	11/12	8/12

The reasons for the similarity of the forecasts, and hence for the representativeness of the overall averages, are several. Forecasters use to a large extent the same data, receive the same news, interact, and draw upon a common pool of knowledge and techniques. The models used often differ substantially, but their outputs are adjusted to reflect the most recent changes in the economy, policies, etc., and it is known that these adjustments reduce the variation among the forecasts (Zarnowitz 1972b, Christ 1975).

Moreover, aggregation over forecasts from business outlook surveys, or other corresponding and contemporary predictions, works to reduce the effects of the outliers (see chapter 15). The aggregate (or average) forecasts are known to be more accurate over time than most of the individual forecasts from the given group.

### 18.2.2 Quarterly Forecasts for the Year Ahead: Cyclical Errors

Table 18.2 presents mean absolute errors and mean errors calculated over a set of quarterly 1-year-ahead forecasts from five widely used sources. The period covered, 1971:2–1985:1, is subdivided in three different ways. Of four equal (14–quarter) subperiods, it is the latest one, 1981:4–1985:1, that shows the largest MAEs. This applies to all four variables included: growth rates in GNP, RGNP, and IPD, and the unemployment rate (UR). But there is no systematic increase in the errors from one period to the next, except for GNP. The mean errors vary in sign and size irregularly across the four periods.

Each of the periods listed in section A of the table includes some especially turbulent times associated with unanticipated turning points in the level of economic activity and rates of growth in output and prices. But the last, 1981:4–1985:1, had the largest share of such events: the severe recession in late 1981 and 1982, the slowdown of mid-1984, and the surprisingly strong disinflation. It is presumably this fact that explains why forecasts for this subperiod were the least accurate.

Indeed, the breakdowns according to cyclical characteristics (sections B and C) disclose much larger and more systematic differences than the division of the period into equal parts. For each variable, the MAEs are much smaller for the predictions relating to business expansions including peaks (44 quarters) than for the predictions relating to business contractions including troughs (12 quarters). Also, the absolute values of the MEs are in each case much smaller for the first than for the second subset of the forecasts (section B).

While the dating of *business cycles* is based on the consensus of the major

**Table 18.2** Some Sources of Variability of Errors in Composite Forecasts of Nominal and Real Growth, Inflation, and the Unemployment Rate, 1971:2–1985:1

Line (1)	Period and No. of Quarters Covered (2)	Medians of Quarterly 1-Year-Ahead Forecasts from Five Sources <sup>a</sup>							
		Mean Absolute Error				Mean Error			
		GNP (3)	RGNP (4)	IP (5)	UR (6)	GNP (7)	RGNP (8)	IPD (9)	UR (10)
<i>A. Four Equal Subperiods<sup>b</sup></i>									
1	71:2–74:3 (14)	1.3	1.1	1.3	0.3	-1.2	0.6	-1.8	-0.1
2	74:4–78:1 (14)	1.4	1.8	1.1	0.8	0.9	1.0	-0.4	-0.5
3	78:2–81:3 (14)	2.3	1.2	1.2	0.5	-2.3	-1.0	-1.1	0.5
4	81:4–85:1 (14)	3.2	2.4	1.6	1.2	2.2	0.6	1.6	-0.3
<i>B. Business Cycle Phases<sup>c</sup></i>									
5	Expansions (44)	1.8	1.2	1.1	0.6	-0.8	-0.4	-0.3	0.1
6	Contractions (12)	3.0	3.0	2.5	1.0	-2.4	2.9	-0.8	-1.0
<i>C. Growth Cycle Phases<sup>d</sup></i>									
7	High-growth (34)	1.7	1.3	0.9	0.6	-0.3	-0.4	0.02	0.03
8	Low-growth (22)	2.7	2.1	2.2	0.8	0.3	1.3	-1.2	-0.3
<i>D. Total Period Covered</i>									
9	1971:2–1985:1 (56)	2.1	1.6	1.4	0.7	-0.1	0.3	-0.5	-0.1

<sup>a</sup>Errors are calculated from the median forecasts by the ASA-NBER survey (ANB), Chase Econometric Associates, Inc. (CHA), Data Resources, Inc. (DRI), Wharton Econometric Forecasting Associates, Inc. (WHM), and Bureau of Economic Analysis of the U.S. Department of Commerce (BEA). On the sources of the forecasts, see notes to tables 18.1 and 18.2; also, Hirsch et al. 1974 (on BEA). For the underlying data, see McNees 1985, table 1, p. 37.

<sup>b</sup>See lines 1–4 col. 2, for the dates of these periods.

<sup>c</sup>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, 1981:4–1982:4. The quarterly dates of peaks and troughs are from the NBER business cycle chronology as used in the BEA monthly publication *Business Condition Digest* (BCD).

<sup>d</sup>High-growth phases are periods during which the mean growth of real GNP exceeded the long-term trend rate (about 3% per year): 1971:2–1972:4, 1975:2–1978:4, 1980:4–1981:1, and 1983:1–1985:1. Low-growth phases are periods of below-trend growth in real GNP: 1973:2–1975:1, 1979:1–1980:3, and 1981:2–1982:4.

turning points in comprehensive economic time series, the dating of *growth cycles* is determined in a similar way from the principal turns in the *detrended values* of such series. What this means in practice here is that periods during which the economy's output grew at an average rate exceeding the long-term trend rate of about 3.3% per year are distinguished from periods during which it grew more slowly. The high-growth phases include recoveries and booms; the low-growth phases include slowdowns and recessions. For the so-defined growth cycles (section C), the contrasts between the phase forecast errors are less sharp than for business cycles but no less regular. The forecasts relating to the high-growth phases (34 quarters) have smaller MAEs than those relating to the low-growth phases (22 quarters), for each variable. The absolute



values of MEs are much smaller for high-growth than low-growth phases in three cases and equal in one (GNP).

To sum up, forecasts of growth in income and output and of inflation and unemployment all tend to be both less accurate and more biased for recessions than for expansions. Similarly, the forecasts for the above-average growth phases look better than those for the below-average growth phases under both criteria. These results are consistent with the earlier ones showing that large errors tend to cluster around business cycle turns, especially peaks (Zarnowitz 1967 and chapter 14, this volume).

The statement just made is not the same as to say that forecasting failures are due to large unanticipated disturbances. Such shocks can and do occur under any economic conditions, yet they seem to cause large errors mainly during slowdowns and contractions. This may be so because it is in these phases, rather than in vigorous recoveries and strong widespread expansions, that the economy is particularly vulnerable. Various stresses and imbalances accumulate gradually as more and more industries approach high capacity operations. Costs of labor, capital goods, and credit typically rise; here and there prices and profits come under squeeze; real shortages appear, growth weakens, and investment begins to decline. Although these internal developments, if permitted to take their course, could alone bring about a downturn, it is also possible for some adverse shocks to speed up this outcome. Yet the same shocks would probably have been weathered by the economy in a less exposed state. The forecaster faces an extremely difficult problem in that (a) it is very difficult to anticipate just when the stresses and imbalances will do their work and (b) the timing of true random shocks that matter is always unpredictable, even if their consequences are not.

In addition, predicting a general downturn is always unpopular, and predicting it prematurely ahead of others may prove quite costly to the forecaster and his customers. On the other hand, most users are likely to await eagerly an upturn during a recognized recession, so forecasts of a recovery will be welcome and often accepted on the basis of early signs of improvement. In this context, it should be recalled that early cyclical indicators had in recent times much longer and more variable leads at peaks than at troughs. (However, their signals of the last recovery came relatively early in 1982, which probably induced some forecasters to err in predicting the recovery too soon.) The peak errors show up during the recession and slowdown periods; the generally smaller trough errors show up during the recovery and speedup periods.

Finally, there is the hypothesis that important macroeconomic functions which are approximately linear as long as there is substantial slack in the economy and relative price stability become nonlinear at high levels of employment and capacity utilization with rising inflation. Econometric models, it is believed, may not be capable of capturing the nonlinearities sufficiently well and hence would perform worse near the peaks of the cycle than at lower levels of macroeconomic activity (see, e.g., Evans 1974, p. 185). This argu-

ment, of course, refers to the endogenous sources of business fluctuations rather than the effects of exogenous disturbances.

### 18.2.3 Quarterly Multiperiod Forecasts from Econometric Services, Business Outlook Surveys, and Time-Series Models

Predicting the developments within the next year or two by quarters is far more difficult than predicting how the economy will fare from year to year. Errors of the forecasts for consecutive quarters typically offset each other to some degree within any year. Also, forecasts for the year ahead can be satisfactory when based on a good record for the first two quarters, and they tend to be more accurate than forecasts with effective spans longer than two quarters (see chapter 14).

However, there is much demand for frequent and detailed predictions, and forecasters have responded by producing quarterly or even monthly forecasts for sequences of 4–8 quarters ahead. The ambitious tendency to disaggregate forecasts over time as well as over space received much support from the falling computation costs in the 1960s and 1970s. With few exceptions, the macroeconomic models now regularly used in commercial forecasting are large, and in several well-known cases, very large.

The great expansion of the models led to expectations of dependably good forecasts, which however met with frequent disappointments. Soon the theoretical basis of the conventional macro models came under sharp attack (Lucas 1976) and some critics proceeded to challenge them with forecasts from vector autoregressive (VAR) models (Sims 1980b).

In econometric forecasting, exogenous variables are projected outside the model, and the model outputs of endogenous variables are as a rule subjected to judgmental adjustments. In contrast, there are no exogenous variables in the VAR models: each of the selected variables is predicted by regression on its own lagged values and those of the others. In the unconstrained model, the only use of economic theory and judgment is in choosing the variables. Since several lags are used for each variable in each equation, the number of variables that the model can accommodate is small. The forecasting process is mechanical and replicable, involving no judgment on the part of the model user.

In practice, this low-cost approach frequently confronts the difficulty of having to estimate many parameters from limited amounts of data with measurement errors. To avoid overfitting and improve forecasts, constraints on the coefficients are imposed in the so-called Bayesian vector autoregressions (BVAR) with the aid of the model builder's prior distributions concerning the stochastic properties of the processes and lags involved (Litterman 1986).

In recent years, the autoregressive integrated moving-average (ARIMA) approach has also been used to forecast selected aggregative variables (Nelson 1972, 1984). Univariate ARIMA models require less simple statistical techniques and computer programs than VAR and more experienced judgment.

They, of course, capture neither the signals nor the noise from the multivariate interactions that are involved in the application of the VAR models.

Table 18.3 draws on recent studies (Lupoletti and Webb 1986; McNees 1986) to compare the performance of these time-series models with several econometric service bureaus and group forecasts from business outlook surveys. To concentrate on the evolution over time, the reported statistics on the root mean square errors (RMSE) are expressed at annual rates and averaged across the forecast horizons for various periods between 1970 and 1985. In addition to these absolute measures of average accuracy (part A), RMSE ratios are used to measure the accuracy of the forecasts relative to that of the corresponding VAR and BVAR projections (part B).

Comparing the periods 1970–75, 1975–80, and 1980–83 (cols. 4–6), one finds rises in the RMSEs for the econometric bureaus' forecasts of GNP, RGNP, and TBR (see lines 1–3, 8–10, and 22–24). But measures from another compilation show the forecasts for GNP and RGNP having smaller average errors in 1980–85 than in 1980–83 (cf. cols. 6 and 10). For inflation, the largest errors are found in 1970–75 and the smallest in 1975–80, but there the differences over time are comparatively small (lines 15–17). Forecasts made within two quarters from business cycle turning points show relatively large RMSEs throughout (col. 9). The large average errors of 1980:4–1983:4 reflect mainly the unexpectedly sharp and long business contraction of 1981–82.

A simple VAR model (see table 18.3, n. *b*) performed generally worse than the econometric services in the early and late 1970s but better in the early 1980s, for both GNP and RGNP.<sup>3</sup> However, the VAR forecasts for inflation were the best in 1970–75, among the best in 1980–83, and the worst in 1975–80. The VAR predictions of interest rates compare poorly with the others in 1970–75 and favorably with others in 1975–80. Around the turning points, and overall, VAR did on the whole not much worse than the econometricians, despite their much more complex and expensive procedures (see the corresponding entries in cols. 3–6 and 9, parts A and B).

The ARIMA and BVAR models also produce mixed results but appear to be more or less competitive with the other forecasters. Judging from the average RMSE ratios, they outperformed VAR in four out of seven cases (cols. 7 and 8).

The last column in table 18.3 sums up some accuracy comparisons of BVAR with econometric and survey forecasts, based on a study by McNees (1986). Litterman has been making BVAR predictions monthly since 1980

3. The VAR model was estimated for the period 1952:2–1969:4, and the obtained coefficients and predictions were then used to forecast each variable for 1970:1–1971:2; this procedure was repeated starting with each successive quarter to produce forecasts with horizons of 1–6 quarters for 1970:1–1983:4. Thus the results are postsample predictions comparable in this respect to the authentic *ex ante* forecasts. However, the data used in the VAR computations were the latest revised estimates available to the authors, whereas the econometric services used of course the preliminary estimates available at the time of the forecast. This could well bias the comparisons in favor of VAR, but there is some evidence that this is not the case (Lupoletti and Webb 1986, table 1 and pp. 267–69).

Table 18.3

## Average Accuracy of Econometric and Other Forecasting Services versus Extrapolations from Time-Series Models, 1970-1985 and Subperiods

A. Measures of Absolute Accuracy									
Root Mean Square Errors (RMSE) <sup>a</sup>									
Line	Forecaster <sup>b</sup>	1970:4 to 1983:4 <sup>c</sup>	1970:4 to 1975:1 <sup>d</sup>	1975:2 to 1980:3	1980:4 to 1983:4	1976:2 to 1982:4 <sup>e</sup>	1980:3 to 1983:4 <sup>f</sup>	Around Turning Points <sup>g</sup>	1980:2 to 1985:1 <sup>h</sup>
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>Growth Rate of Gross National Product (GNP)</i>									
1	Chase	4.0	2.3	4.0	5.4			4.4	4.3
2	DRI	3.5	2.3	3.0	5.0			4.0	3.9
3	Wharton	3.6	2.7	3.0	4.9			4.2	4.2
4	ANB								3.8
5	VAR	4.3	4.9	3.8	4.3	4.6	4.5	4.1	
6	ARIMA					4.8			
7	BVAR						5.0		4.3
<i>Growth Rate of Real GNP (RGNP)</i>									
8	Chase	3.0	2.4	2.8	3.8			3.4	3.0
9	DRI	3.0	3.0	2.5	3.6			3.6	2.8
10	Wharton	2.8	2.7	2.5	3.2			3.3	2.7
11	ANB								3.0
12	VAR	3.6	4.3	3.2	3.4	3.5	3.5	3.9	
13	ARIMA					3.4			
14	BVAR						2.6		2.3
<i>Rate of Inflation in the GNP Implicit Price Deflator (IPD)</i>									
15	Chase	2.6	2.6	1.7	2.0			2.3	1.9
16	DRI	2.1	2.8	1.6	1.8			2.3	1.8
17	Wharton	2.0	2.4	1.6	1.9			2.0	2.0
18	ANB								1.4
19	VAR	2.2	2.0	2.3	1.8	2.0	2.4	2.5	
20	ARIMA					1.8			
21	BVAR						3.5		3.3
<i>90-Day Treasury Bill Rate (TBR)</i>									
22	Chase	2.5	1.6	2.4	3.3			2.7	2.6
23	DRI	2.3	1.5	1.9	3.4			2.6	3.0
24	Wharton	n.a.	n.a.	2.2	3.0			n.a.	3.0
25	VAR	2.4	2.1	1.8	3.6		3.6	2.8	
26	BVAR						3.3		3.2
B. Measures of Relative Accuracy									
RMSE Ratios (RMSE - VAR = 100) <sup>i</sup>									
RMSE - BVAR = 100 <sup>j</sup>									
Line	Forecaster	1970:4 to 1983:4	1970:4 to 1975:1	1975:2 to 1980:3	1980:4 to 1983:4	1976:2 to 1982:4	1980:3 to 1983:4	Around Turning Points	1980:2 to 1985:1
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>GNP</i>									
27	Chase	0.93	0.47	1.05	1.26			1.07	1.00
28	DRI	0.81	0.47	0.79	1.16			0.98	0.91
29	Wharton	0.84	0.55	0.79	1.14			1.02	0.98

(continued)

Table 18.3 Continued

		B. Measures of Relative Accuracy							
		RMSE Ratios (RMSE - VAR = 100) <sup>c</sup>							RMSE -
		1970:4	1970:4	1975:2	1980:4	1976:2	1980:3	Around	RMSE -
Line	Forecaster	to	to	to	to	to	to	Turning	to
(1)	(2)	1983:4	1975:1	1980:3	1983:4	1982:4	1983:4	Points	1980:2
		(3)	(4)	(5)	(6)	(7)	(8)	(9)	1985:1
									(10)
<i>GNP</i>									
30	ANB								0.93
31	ARIMA					1.04			
32	BVAR						1.11		
<i>RGNP</i>									
33	Chase	0.83	0.56	0.88	1.12			0.87	1.30
34	DRI	0.83	0.70	0.78	1.06			0.92	1.22
35	Wharton	0.78	0.63	0.78	0.94			0.85	1.17
36	ANB								1.25
17	ARIMA					0.97			
38	BVAR						0.74		
<i>IPD</i>									
39	Chase	1.18	1.30	0.74	1.11			0.92	0.58
40	DRI	0.95	1.40	0.70	1.00			0.92	0.55
41	Wharton	0.91	1.20	0.70	1.06			0.80	0.61
42	ANB								0.50
43	ARIMA					0.90			
44	BVAR						1.46		
<i>TBR</i>									
45	Chase	1.04	0.76	1.33	0.92			0.96	0.81
46	DRI	1.05	0.71	1.06	0.94			0.93	0.93
47	Wharton	n.a.	n.a.	1.22	0.83			n.a.	
48	BVAR						0.92		

Sources: Cols. 3-9: Lupoletti and Webb 1986, tables 2-8; col. 10: McNees 1986, tables 1, 2, 5, and 6.

<sup>a</sup>Averaged across forecast horizons as follows: cols. 3-6, 8, and 9, means of RMSE for horizons of 1, 2, 4, and 6 quarters; col. 7, means for horizons of 1, 2, and 4 quarters; col. 10, means for horizons of 1-8 quarters.

<sup>b</sup>VAR = unrestricted vector autoregressive model with six lags for each of five variables (percentage changes in the monetary base, RGNP, and IPD; the manufacturing-capacity utilization rate; and the 90-day TBR); see Webb 1984 and Lupoletti and Webb 1986. ARIMA = univariate autoregressive integrated moving-average model (Nelson 1972, 1984). BVAR = Bayesian vector autoregressive model with six lags for each of seven variables (annual growth rates of RGNP and IPD; UR; lagged levels of the money supply MI and of gross private domestic investment; 4-6 month commercial paper rate and the change in business inventories); see Litterman 1986. The BVAR model used in 1980-1983 (col. 7) consists of six variables (it does not contain the inventory series and uses real business fixed investment instead of GDP). On the sources of the other forecasts, see references in notes to tables 18.1 and 18.2. All underlying data are authentic (postsample) forecasts measured at annual rates.

<sup>c</sup>The dates are for 1-quarter forecasts; 2-quarter forecasts: 1971:1-1983:4; 4-quarter forecasts: 1971:3-1983:4; 6-quarter forecasts: 1972:1-1983:4.

<sup>d</sup>The dates are for 1-quarter forecasts; see n. c on the starting dates for 2-, 4-, and 6-quarter forecasts.

<sup>e</sup>The dates are for 1-quarter forecasts; 2-quarter forecasts: 1976:3-1982:4; 4-quarter forecasts: 1977:1-1982:4.

**Table 18.3** Continued

<sup>f</sup>The dates are for 1-quarter forecasts; 2-quarter forecasts: 1980:4–1983:4; 4-quarter forecasts: 1981:2–1983:4; 6-quarter forecasts: 1981:4–1983:4.

<sup>g</sup>Covers forecasts made within 2 quarters from a business cycle turning point (NBER dates). For periods covered, see n. c.

<sup>h</sup>Dates show the forecast period covered. The RMSE entries are in percentage points, cumulative growth at annual rates (GNP, RGNP, IPD) and in percentage points, cumulative changes (TBR). Chase, DRI, and Wharton are “early-quarter” forecasts; ANB are “midquarter” forecasts. BVAR are based on data available early in each quarter. The BVAR predictions based on data as of midquarter and comparable to ANB are GNP, 4.1; RGNP, 2.4; IPD, 2.8.

<sup>i</sup>Based on entries in part A, cols. 3–9. Ratios of RMSE of other forecasts to the corresponding RMSE of VAR.

<sup>j</sup>Based on entries in part A, col. 10. Ratios of RMSE of other forecasts to the corresponding RMSE of BVAR. The BVAR figures used in the ratios for Chase, DRI, and Wharton are those in lines 7, 14, 21, and 26. For the BVAR figures used in the ratios for ANB, see n. h.

(see table 18.3, n. b). Real growth was predicted much better by his model than by the other forecasters in 1980–85, but inflation was predicted much worse.

With respect to differences by forecast horizon (which are ignored in table 18.3), pairwise comparisons of the RMSEs for each VAR with each model suggest that the relative performance of VAR improved with the length of forecast. Cases in which VAR had smaller RMSEs than the other forecasts account for 24%, 25%, 42%, and 42% of all comparisons for horizons of one, two, four, and six quarters, respectively. Pairwise comparisons with BVAR show that each of the forecasters included in table 18.3 had smaller RMSEs for IPD and TRB at all horizons. The reverse obtains for RGNP, where BVAR produced the best results in half of the shortest and all of the longer forecasts. For GNP, BVAR was worst in each case over horizons of 1–4 quarters but better than the others in most of the comparisons for horizons of 6–8 quarters. When more variables and more forecasters are included, BVAR comes out ahead in most of the comparisons for the period 1980:2–1985:1 (see Granger’s comment on McNees 1986). Fragmentary results for the more recent years, however, suggest some deterioration. Thus Litterman’s BVAR forecasts for 1984:4–1985:4 and 1985:4–1986:4 predicted high growth rates in real GNP and such of its components as durable-goods consumption, gross private domestic investment, and residential construction. The corresponding realizations were much lower in 1985–86. Most forecasters have been much less optimistic than BVAR, and more accurate.<sup>4</sup>

4. The following tabulation compares some of the forecasts from Litterman 1984 with the actual percentage changes for 1985 (fourth quarter over fourth quarter):

	RGNP	C	CD	CNS	GPDI	NRFI	RFI	IPD
BVAR	3.7	4.6	10.0	3.6	6.2	6.1	12.3	3.2
Actual	2.9	3.5	6.2	3.1	0.5	6.6	7.8	3.3

### 18.3 Conclusion, Implications, and Further Thoughts

There is no evidence, here or elsewhere, that macroeconomic forecasts in the United States have grown systematically worse, that is, less accurate, more biased, or both. Rather this paper argues and to some extent documents that the failures of forecasting are related to the incidence of slowdowns and contractions in general economic activity. Not only the forecasts of real GNP growth and unemployment but also those of nominal GNP growth and inflation tend to go seriously wrong when such setbacks occur. This result seems strong, though qualitatively not surprising: it confirms and extends earlier indications of typically large turning-point errors.

The question that naturally arises at this point is, do such findings have useful lessons for producers and users of macroeconomic forecasts? It is clear that forecasters cannot afford to wait for long expansions to prove the usefulness of their own activities. Instead, as our results demonstrate, there is urgent need for the forecast makers to increase their ability to anticipate the retardations and declines in aggregate demand, output, and employment—and for forecast users to pay particular attention to such efforts and reward any resulting successes.

To be sure, all this is much easier said than done, but it seems highly probable that economic forecasting can be improved to some degree and that we are far from having reached the limits of this process. After all, macroforecasting as an explicit activity put to practical uses, yielding recorded and testable results, and subject to the disciplines of market and research is very young indeed. New and useful insights will not come easy here but will be achieved, and new methods and new applications of old methods are being developed continually.

The four active, broad approaches to short-term forecasting of the economy at large are time-series models, econometric models, anticipations surveys, and cyclical indicators. Each of these corresponds to a particular aspect of the entire task. Thus, time-series models are best equipped to exploit intensively the information contained in the past history of the single or several series to be predicted; macroeconometric models, to quantify the predominant relationships that the theory suggests exist among a larger (but not overly large) number of variables; anticipation surveys, to estimate aggregates of plans or intentions of economic agents for variables over which these agents exercise considerable control; and the indicators, to signal and confirm certain recurrent business cycle events. These are distinct but interrelated functions. In the present practice, none of them is performed very well because of paucity of generally agreed upon and successfully tested economic theories that would

---

C, CD, and CNS denote the percentage changes in real consumption for total, durables, and nondurables and services. GPD, NFRI, and RFI denote percentage changes in total (gross private domestic), nonresidential fixed, and residential fixed investment.

provide strict guidance for macromodeling and because of inadequacies of the available data, estimation, and surveying techniques. Yet there are significant advantages to using each class of models or methods for the task to which it is best suited. In short, contrary to some partisan assertions and criticisms, the four approaches are essentially complements, not competitors or substitutes. They need to be refined and used in combination so as to contribute to the improvement of the forecasts.

To illustrate, the blending of time-series analytic and traditional econometric methods can result in better selection and projection of exogenous variables. The devices to be used for this purpose are tests for exogeneity and extrapolations based on full information contained in the past history of the series to be predicted and related forward-looking data (e.g., for federal government purchases, recent congressional appropriations and debates). Further, time-series models can be constructed for, and applied to, the residual errors from econometric equations as these terms are often far from being purely random. The transfer functions which thus combine regression with time-series models would be expected to have greater predictive power than either type of model alone.<sup>5</sup>

Similarly, consistency with the lessons from anticipated surveys and sequences of leading, coincident, and lagging indicators should enhance the usefulness of any macroforecasting model. Probably the best way to achieve this objective is to include the relationships involving the principal survey data and indicators directly in the model. To mention just one important area, promising because of strong elements of executive planning and long gestation periods, forecasting business expenditures for new plant and equipment should draw on surveys of backlogged and newly approved capital appropriations, surveys of anticipated expenditures, construction contracts for plant buildings, and new orders of nondefense capital goods industries.

The long record of leading indicators in predicting business cycle turning points is encouraging. With the aid of suitable time-series analytic transformations and decision rules, it should be possible to reach considerably better results yet in this respect. The main practical problem for this approach lies in false signals; the errors of the other type—missed turns—are rare and relatively unimportant. To reduce the risk of false warnings, a system of sequential signals from both leading and confirming indicators has been proposed and tested with generally positive results (Zarnowitz and Moore 1982 with an update in Moore 1983; Niemira 1983). Predictions with the composite index or vector of selected leading indicators can improve on autoregressive forecasts of changes in real GNP, industrial production, and the rate of unemploy-

5. On time-series analysis and econometric models. see Zellner and Palm 1974; on transfer functions, with applications. Pindyck and Rubinfeld 1981; on exogeneity tests, Granger 1969 and Sims 1972. On further developments and the more radical and controversial "index models" (using restricted VARs for business cycle analysis). see the collection of papers and comments in Sims 1977.



ment (Vaccara and Zarnowitz 1978; Auerbach 1982). Signals of cyclical downturns from the leading index can take into account estimated probability distributions of phase durations and percentage changes in the index (Neftçi 1982, Palash and Radecki 1985).

Of particular interest is a technique which combines time-series models with Monte Carlo simulations to generate repeated sample paths of the predicted series and probability distributions over the relevant turning points (Wecker 1979). This analysis has recently been extended to multivariate models for related indicator series, with explicit assessments of uncertainty in the estimates and of turning-point probabilities (Kling 1986).

To conclude, forecasters tend to rely heavily on the persistence of trends in spending, output, and the price level. To the extent that inertia prevails in the economy's movement, their predictions turn out to be roughly right, at least directionally, most of the time. But the inertia, although helpful in this sense, is only a part of the story, and such forecasts suffer from missing business cycle turns and underestimating recessions and recoveries with respect to both their real and nominal effects. These errors are only in part due to the impact of the many inevitable random disturbances to the economy that cannot be anticipated. Although variable in their observed durations and amplitudes, the expansions and contractions in the major economic aggregates, both in levels and deviations from long-term trends, show many important recurrent features. These regularities, as reflected in the relative movements of cyclical indicators, should and can be better captured in the work of macroeconomic forecasters.