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: On Short-Term Predictions of General Economic Conditions

Chapter Author: Victor Zarnowitz

Chapter URL: http://www.nber.org/chapters/c10384

Chapter pages in book: (p. 383 - 413)

IV Forecasting

13 On Short-Term Predictions of General Economic Conditions

Part IV (chapters 13–18) is about forecasting the course of the economy in the near future, that is, up to 2 years, or eight quarters, ahead. Only the true ex ante predictions that are nontrivial and verifiable are considered, and only for a relatively small number of principal U.S. macroeconomic variables, real and nominal. The period covered is limited to the last three or four decades by the availability of the forecasts, but the spectrum of sources and methods covered is broad. Empirical studies of how macroeconomic forecasts and expectations are formed and what their properties and accuracy are have useful lessons despite their relatively short history. This chapter will serve as an introduction and guide to the subject; the chapters that follow deal much more thoroughly with several of its dimensions.[1]

13.1 Functions and Contents

Economic forecasts refer to economic aspects of unknown events, whether in the past, present, or future. It is of course the future that attracts most attention, since the practical purpose of the forecasts is to help formulate and improve public and private plans and decisions, which are necessarily forward looking. Formal forecasts of economists attempt to take into account the collective effects of these decisions, but they also influence in various ways and degrees the informal expectations of consumers, investors, business managers, and government officials.

The "unknown events" come in many different types and forms. The most common and regular targets of macroeconomic forecasting are the rates of growth in real GNP and its major expenditure components, the unemployment

1. Some of sections below draw in part on material in my earlier publications (Zarnowitz 1972b, 1974) and update it.

rate, and inflation and interest rates. Many professional forecasts are more detailed and extend to such cyclically sensitive series as corporate profits, industrial production, and housing starts. Large econometric model services help predict hundreds of variables that are of interest to subscribers. Financial forecasters concentrate on stock and bond prices and yields and on exchange rates, all of which are highly volatile, responsive to a continuous flow of news and rumors that affect traders' expectations, and notoriously difficult to predict.

Indeed, it is often argued that financial and other major economic series behave largely as random walks and hence cannot be predicted at all. But the overall stock market movements over intermediate and longer horizons, for example, have clearly strong trend-cycle components. They are both persistent and pervasive (large majorities of individual stock prices participate), anticipatory of business cycle and growth cycle turns, and related to other highly cyclical variables, notably corporate profits and (inversely) interest rates.[2]

Predicting routine events in economics and business generally takes the form of periodic forecasts of time-series values. In addition, there are events of special concern such as the turning points of business and growth cycles, financial panics and major bankruptcies, strikes, international crises and wars, and so on. Forecasters may or may not try to anticipate cyclical turns with the aid of leading indicators, but in any event their time-series predictions have implications for the timing of these events that need to be monitored and checked. Exogenous shocks, such as a sharp rise in oil prices due to a foreign cartel action or a threat of war, are not amenable to economic forecasting but their consequences for general business activity are.

Real aggregates and price indexes typically have upward trends reflecting economic growth and inflation, respectively; nominal aggregates expand for both reasons. Growth trends are long-term movements that dominate comprehensive output, employment, and related income and expenditure series in real terms across decades. In the period covered by our forecasts, inflation too has been persistent as well as pervasive. Cyclical fluctuations prevail over intermediate horizons of several quarters and years. In the shortest run, measured in weeks and months, changes in most economic time series are heavily influenced by irregular variations from random causes and by intrayear, approximately periodic, seasonal movements. 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. In practice, then, the systematic part of the time series covered by the short-term macroeconomic forecasts (as defined above) consists of elements of trend and, to a larger extent, cyclical movements.

It follows that a forecast should be judged successful if it approximates well that systematic part of its target (which includes the effects of past shocks and

2. See chapter 11, sec. 11.4.2, for more detail and references.

seasonal innovations to the extent they are knowable). But this task is generally much more difficult than it sounds. Movements that are "systematic" are persistent or recurrent, but they are by no means predetermined, periodic, or repetitious. As noted earlier, trends and cycles 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 that there are few (if any) constant quantitative rules (e.g., time-invariant linear econometric equations) to help the macroeconomic forecaster effectively and consistently over more than a few years or from one business cycle to another.

13.2 Sources and Uses

Business and economic forecasting in the United States today is a highly diversified "industry" of significant size. A large majority of its members are business economists whose main function is to provide information to improve managerial decisions. The membership of the National Association of Business Economists grew rapidly from 322 in 1959 (the year NABE was founded) to 1,682 in 1969 and 2,749 in 1979; it peaked at 3,491 in 1983, declined to 3,098 in 1987, and rose to more than 3,300 in 1989 and 1990. By far most of these people are professionals working for private companies in manufacturing, finance, trade, services, consulting, etc.; some are in government at all levels and in academic institutions. They use macroeconomic forecasts in their work as inputs to assessments of prospects for their own activities or concerns (firms, industries, regions), but only a minority are regular producers of such forecasts. Thus the "macro panel" of NABE (which quarterly updates predictions of annual changes in real GNP and components, inflation and interest rates, and other aggregative variables) now has 125 participants, and the actual number of regular macroeconomic forecasters is probably of the order of 200–250.[3]

The forecasting units vary in size from individuals and small teams to sizable economic divisions of some large corporations and multibranch specialized consulting and forecasting firms. Some of the last operate large-scale econometric models and provide customer services internationally.[4] The demand for macroeconomic forecasts is to a substantial degree met by subscriptions to such services and also to publications that frequently survey groups of professional forecasters and list their individual and average predictions.[5]

3. I am very grateful to David L. Williams, secretary-treasurer of NABE, for help in collecting the factual information used in the text paragraph above. The membership figures exclude student and institutional members.

4. Among the largest and best known are Wharton Econometric Forecasting Associates, Inc. (WEFA), and Data Resources, Inc. (DRI). See Adams 1986 and Eckstein 1983.

5. Monthly surveys of forecasters available by subscription are the *Blue Chip Economic Indicators* (since 1978) and *Economic Forecasts: A Worldwide Survey* (since 1984). A quarterly survey in the public domain was conducted by NBER and the American Statistical Association from

Since the supply comes from a small number of relatively large producers and many small ones, the forecasting industry can be described as a mixture of oligopolistic and competitive elements. But it is also arguable that the market for forecasts is one of monopolistic competition as the overall number of sellers is large, the products are differentiated in several respects, and the barriers to entry seem low. In other countries, macroeconomic forecasting is generally much more concentrated, either in a few private sources or in government agencies and publicly supported economic research organizations.[6]

U.S. government forecasts that are designed to serve as inputs into the economic policy-making originate in several agencies: the Council of Economic Advisers (CEA), the Office of Management and Budget (OMB), the Treasury, Federal Reserve Board (FRB), and the Bureau of Economic Analysis (BEA) in the Department of Commerce. Many of these forecasts are for internal uses only and based on economists' models and judgments (FRB and BEA have their own macroeconometric models). The official forecasts that are published, such as those of the CEA in the *Economic Report of the President* (January or February of each year), are end products of interaction among top policymakers as well as their economic advisers.

Those who must forecast regularly and frequently are likely to be absorbed by technical requirements of their profession: monitoring and processing information, analyzing current economic and political developments, preparing reports, and interpreting results. Most are pragmatic and use all data and approaches that they deem helpful to improve their predictions; few spend much time on working with and testing specific theoretical models. The principal scientific and academic use of the forecasts is to test the various hypotheses, models, and methods employed by the forecaster, but this task is largely left to the outside observer. The same applies to another, quite different but also important aspect of predictions of economic change, namely, that they may be useful in providing data to study how expectations are formed, transmitted, and revised.

All these are definitely bona fide uses of forecasts. But some forecasts are at least occasionally and secondarily used as means of communicating intentions and influencing opinion, which may bias them or make them otherwise questionable.

In short, economic and business forecasters serve many different masters.

1968:4 to 1990:1. It is being continued, in essentially unchanged form, by the Federal Reserve Bank of Philadelphia.

6. The former case is well represented by Canada: see the instructive analysis in Daub 1987, ch. 6. The latter situation prevails in several European countries. For example, in Germany six economic research institutes dominate the field; elsewhere the most influential forecasts originate in ministries of finance and economic affairs, central banks, universities, and research institutes. There are probably many private business forecasts as well but few are collected and publicized. For a comprehensive index of organizations engaged in macroeconomic forecasting worldwide, see Cyriax 1981.

Much of their output is communication to the business or government administrators who employ their services; some is being done for the outside world: peers, professionals, the interested public. And it is not unusual for their products to have both internal and external uses. For example, an econometric model developed at a university may serve as a basis for scientific work and also as a source of forecasts circulated to business subscribers; aggregate forecasts by the economic staff of a corporation are a basic input to micropredictions of sales, etc., but are also used in the company's publications and speeches by its executives; and so on.

Forecasters face all sorts of conflicts of evidence and opinion, which they often resolve by various internal compromises; since their work is essentially conjectural, much of this seems inevitable. The multipurpose nature of some forecasts may complicate the situation considerably by bringing forth some conflicts of interest as well.

13.3 Quality and Accuracy

The decision maker who knows the cost of acquiring and using the forecast and the returns attributable to it should in principle be able to evaluate the quality of the forecast exactly, at least in retrospect. This is an ideal condition but good approximations to it may exist in massive routine applications of simple methods to replicable problems. Thus a manufacturing company with thousands of products must use low-cost time-series models for purposes of inventory control and production scheduling; it can experiment with alternative models and choose the most cost-effective ones with considerable confidence. However, economic forecasting generally cannot be reduced to such situations. In macroforecasting especially, there is a wide choice of sophisticated models and techniques, a major role for judgment, but little opportunity for any controlled experimentation. Microlevel information required for a comparative analysis of costs and returns is usually confidential and not available to an outside observer; on a macrolevel, such information may not exist at all or be very underdeveloped and costly.

However, with reasonably reliable and prompt information, it is possible to assess short-term aggregate forecasts in a meaningful way by ex post comparisons with actual outcomes. Verifiability so defined is a necessary condition for the forecasts to be potentially useful. Hence it seems natural to view the overall *accuracy* of a given set of predictions as the principal single aspect of their *quality*, that is, goodness in use. But accuracy is relative and it depends on other characteristics that differentiate forecasts such as scope, span, and timeliness. These too codetermine quality and are in principle amenable to objective measurement.

Comparisons of accuracy are often impeded because the differences among forecasts with respect to their other attributes are difficult to allow for by standardization or classification. The lack of information about the costs of forecasts may seem to be a principal obstacle. But cost levels and differentials have been greatly reduced by computer technology and competition among forecast makers and collectors. Although surely relevant, they are probably no longer of major importance for ranking macroeconomic predictions produced by different sources and methods.

Information about the models or the reasoning behind the forecasts is clearly desirable in general and critical for some purposes of research and knowledge. Many business forecasts come with general explanations of the underlying assumptions and arguments, and econometric model services sell large amounts of numerical detail on the equations and adjustments used in their control solutions, alternative simulations, etc. But it is much easier to assess the accuracy of forecasts than the quality of economic analysis and judgment, and the information necessary for the latter task is often neither provided to nor demanded by commercial and lay users.

It is true that a prediction can be "correct for the wrong reasons"—although based on assumptions contrary to fact, it may fortuitously produce only a small error. The converse is also possible, as when a model supported by theory and past experience fails because of some unanticipated shock. This suggests that a quantitative analysis of forecast errors is not sufficient; a qualitative analysis of how each prediction was derived is needed to evaluate the forecasts. However, this argument can be carried too far. Individual predictions may indeed suffer from excusable assumptions about "exogenous," perhaps noneconomic, events, but if a forecaster's performance is below par on the average over time, it is hard to accept wrong assumptions as a justification.[7]

More generally, a few sporadic successes or failures do not prove that a given source or method of forecasts is or is not accurate. To reduce the role of chance, measures of average accuracy are needed, and they are the more informative the longer and more varied the periods covered. Unfortunately, the available samples of forecasts are mostly small, since few forecasters have produced long, consistent time series of verifiable predictions.[8]

In the end, knowing the size of prediction errors is necessary for any appraisal of the consequences of these errors, although it is usually not sufficient. In choosing the products of different forecasters, users act rationally when they prefer those with comparative advantages in past accuracy or at least attach to them greater weights.[9]

7. It is generally prudent for the forecaster to state carefully his or her basic assumptions, and the information may help the forecast users and judges. However, surrounding predictions with hedges against all kinds of unforeseeable events detracts from their usefulness, particularly for business purposes.

8. The samples are numerous and diverse in terms of sources, methods, variables, and predictive horizons, but this "cross-sectional" richness is not a good substitute for the paucity of long time series of forecasts.

9. Whether superior forecasters exist is a related but separate matter, about which more later in sections 13.8 and 13.9 and chapter 15.

13.4 Summary Measures of Absolute Accuracy

The choice between different measures of accuracy depends upon the forecaster's or user's conception of how errors of different types and sizes reduce the usefulness of the forecast. For example, if the loss depends simply on the size of the difference between the predicted and actual value, the appropriate summary measure is the mean absolute error (MAE). If large errors in either direction are considered much more serious than the small ones, squares of the differences should be used, and the proper average is the root mean square error (RMSE).[10]

In still different situations, the sign as well as the size of the error may matter so that the loss function is asymmetric, with underestimates preferred to overestimates, or vice versa. For example, if management would rather err on the side of too low than too high inventory holdings, it may prefer underestimates of sales, that is, penalize them less than overestimates in weighting. However, it does not follow that such a user should favor forecasts with a built-in corresponding "bias": it seems best for the forecaster to produce best unbiased predictions and for the user to apply to the results whatever his or her loss function is, by appropriately weighting the errors. One can conceive of an indefinite number of diverse individual loss functions; for example, turning-point errors may be treated as particularly serious, errors smaller than some present threshold values may be ignored as implying a zero loss, and so on. But little is known about the distribution of user performance, and for practical reasons of general acceptability and comparability, only a few simple types of error measures are commonly in use (mostly MAE and RMSE).

Table 13.1 covers a large number of predominantly judgmental forecasts by business economists and some others, summed up in two sets; the principal government forecast; and the longest series of forecasts made by econometricians working with a family of macroeconomic models. The group averages from surveys of forecasters conceal the dispersion of errors in the predictions of individual participants, which is often large, and they are always more accurate over time than most of their components (see chapters 15 and 16). For the end-of-year predictions of nominal GNP growth in the year ahead, the MAEs show these broad "consensus" forecasts to be about as accurate as the government and econometric forecasts (cf. cols. 2 and 6 with 4 and 8, lines 2–4). These measures refer to periods between 1963 and 1989, range from 0.8 to 1.3 percentage points, average 1.0, and have a standard deviation of 0.2. The earliest available collection of forecasts, for 1956–63, shows a larger MAE of 1.6 percentage points (line 1).

Similarly, the comparisons across the different sets of annual forecasts of

10. The RMSE is a particularly convenient measure mathematically and statistically because it is optimal under a quadratic loss criterion, corresponds to the ordinary least squares estimation procedure, and lends itself to decomposition into systematic (bias, inefficiency) and residual variance components. The MAE corresponds to an alternative V-shaped loss function, where the cost of (loss due to) error depends linearly on the absolute size of the error.

Table 13.1 **Annual Forecasts of Percentage Changes in Nominal and Real GNP and IPD: Mean Absolute Errors and Mean Errors, 1956–89**

Line	Period Covered	Private Judgmental Forecasts, Mean[a]		CEA		ASA-NBER, Median		Michigan Model (RSQE)[b]		Preliminary Data[c]		Mean Absolute % Change[d]
		MAE	ME	MAE	ME	MAE	ME	MAE	ME	MAE	ME	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
						Gross National Product (GNP)						
1	1956–63	1.6	−0.4							0.5	−0.3	5.0
2	1963–76	1.0	−0.7	0.9	−0.2			1.3	−0.5	0.3	−0.3	7.9
3	1969–76	0.8	−0.4	0.8	0.2	1.0	−0.3	1.0	−0.1	0.3	−0.3	8.4
4	1969–89			1.0	0.2	1.2	−0.1			0.5	−0.5	8.4
						GNP in Constant Dollars (RGNP)						
5	1959–67	1.3	−0.9					1.0	−0.5	0.5	−0.2	4.3
6	1962–76			1.1	0.6			1.4	0.2	0.4	−0.1	4.1
7	1969–76			1.2	0.8	1.0	0.7	1.6	0.8	0.3	0.1	3.6
8	1969–89			1.0	−0.3	1.1	−0.1			0.6	0.0	3.2
						GNP Implicit Price Deflator (IPD)						
9	1959–67	0.6	0.2					0.7	0	0.3	−0.3	1.9
10	1962–76			1.0	−0.5			1.0	−0.5	0.4	−0.3	4.2
11	1969–76			1.4	−0.6	1.3	−0.9	1.4	−0.9	0.4	−0.4	5.9
12	1969–89			1.0	0.03	1.2	−0.2			0.6	−0.4	5.6

Sources: Lines 1–3, 5–7, and 9–11: based on tables 14.1, 14.2, and 14.3 (see chapter 14 for detail). Lines 4, 8, and 12: based on author's files and calculations; see also Moore 1983, tables 26.3 and 26.4; *Economic Report of the President;* and *Budget of the U.S. Government.*

Note: MAE (mean absolute error) $= 1/n\ \Sigma\ |E_t|$, where $E_t = P_t - A_t$; P_t is the predicted value and A_t is the actual value (first estimate); and Σ denotes summation over all n years covered. ME (mean error) $= 1/n\ \Sigma\ E_t$. All measures refer to annual percentage changes and are in percentage points.

[a]Line 1: forecasts from (1) *Fortune* magazine; (2) Harris Bank; (3) IBM; (4) National Securities and Research Corporation, (5) NICB Economic Forum; (6) School of Business Administration, University of Missouri; (7) Prudential Insurance Company of America; (8) University of California in Los Angeles (UCLA) Business Forecasting Project; (9) Livingston Survey, mean; and (10) N.Y. Forecasters Club, mean. Lines 5 and 9: sources 2, 4, 5, and 8.

[b]Research Seminar in Quantitative Economics of the University of Michigan.

[c]Based on the first official estimates following the year for which the forecast was made as compared with the revised data from U.S. Department of Commerce, BEA, 1977; SCB, 1991.

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

real GNP growth (RGNP) show fairly small and unsystematic differences in overall accuracy for each of the four periods covered between 1959 and 1989. Here the MAEs have a range of 1.0–1.6 percentage points, mean of 1.2, and standard deviation of 0.2 (lines 5–8).

For the corresponding forecasts of inflation in terms of the implicit price deflator (IPD), the MAEs have a wider range of 0.6–1.4 percentage points, a mean of 1.1, and standard deviation of 0.3. However, the interforecast differences in each line are particularly small here (lines 9–12).

The errors in table 13.1 are measured from preliminary data first published after the end of the forecast target year. This allows us to compare the arithmetic and absolute means of these forecasting errors with the corresponding averages for the measurement errors defined as differences between the percentage changes in the preliminary and revised data. The MAEs of the preliminary data, so computed, have a range of 0.3–0.6 and average 0.4 (col. 10); they are quite sizable compared with the MAEs of the forecasts proper, which have a range of 0.6–1.6 and average 1.1. (Compare these measures also with the mean absolute percentage changes in preliminary data, which interestingly increased over the successive periods for IPD and GNP but decreased for RGNP; see col. 12.) The early data tended to underestimate the revised data by about −0.3 (col. 11).

Predictions of the rates of change in GNP and RGNP made *quarterly* for the year ahead show similar MAEs of 1.1–1.4 even for most subperiods of the difficult decade of the 1970s, but they show much higher MAEs of 1.8–3.2 percentage points for some recent intervals (1978–85 for nominal, 1974–78 and 1981–85 for real growth; see table 18.2 below). The corresponding measures for IPD inflation fall in the 1.1–1.6 range. The worst errors coincide with major cyclical changes such as the recessions of the mid-1970s and early 1980s and the disinflation of 1981–85.

The absolute or squared errors increase systematically with the span of forecast in quarters. For example, the MAEs of the GNP growth predictions in 1970–75 fall in the ranges of 0.4–0.5, 0.8–1.0, 1.7–1.9, and 2.8–3.5 for one, two, four, and eight quarters ahead, respectively (see table 14.5).

13.5 When and Why Growth and Inflation Were Underpredicted or Overpredicted

In the 1950s and 1960s forecasters generally underpredicted the nominal GNP growth in years of cyclical expansion—that is, most of the time. This implies negative averages of errors measured as differences, predicted minus actual values. Thus, the errors of eight sets of GNP forecasts for 1953–70, in billions of dollars, had means of −2.5 for the base (current-year) levels, −4.0 for the base-to-target changes, and −6.5 for the target (next-year) levels.[11] Positive errors (i.e., overestimates of GNP changes or levels) occurred

11. The level errors are equal to the algebraic sums of the corresponding base and change errors. For the sources of the eight forecasts, see table 13.1, n. *a*.

in this period in only six years, each of which was associated with a recession or a major slowdown (1954, 1958, 1960, 1962, 1967, and 1970). The largest underestimation errors occurred in times of booming economic activity, later accompanied also by rising inflation (1955, 1965–66, and 1968–69) (for details see Zarnowitz 1974, table 5 and pp. 578–80).

Thus the underestimation of GNP changes probably reflected in the main an underestimation of real growth. The few year-to-year decreases were more often missed (i.e., increases were predicted instead) than either under- or overstated. RGNP forecasts are not available for the early postwar period, but predictions of the index of industrial production (IP) are and they provide some confirming evidence. Of the seven sets of IP forecasts between 1951 and 1963, all but one have negative mean errors of changes and all but two have negative mean errors of levels (Zarnowitz 1967, table 4, p. 34).[12]

Scattered annual forecasts of the consumer and wholesale price indexes in 1949–67, assembled and examined in Zarnowitz 1969b, show no evidence of an overall bias for CPI and a prevalence of overestimates for WPI. Forecasters tended to underestimate the large changes and to overestimate the small changes in the indexes, and most of the price changes in this period were relatively small. There is much extrapolative inertia in these forecasts, although most of them were more accurate than naive model projections and produced definite positive correlations between the predicted and actual index changes. This evidence too is consistent with the hypothesis that it was mainly real growth that was underpredicted in the GNP forecasts summarized here.

The overall timidity of these predictions can be understood only in their historical context. Forecasters used the available data, which referred mainly to the interwar period, while dealing with the very different economic environment of the postwar era. Many expected a replay of what happened after World War I, repeatedly using analogies that turned out to be basically false. The recession of 1948–49 occurred later and was less severe than they had projected. Before the proper lessons from the errors were drawn, the shocks of the Korean War created new uncertainties. The recessions that followed in 1953–60 were relatively mild but discouragingly frequent (three in 8 years).

However, forecasters gradually learned to expect higher rates of economic growth. RGNP increases were strongly underestimated in 1959–67 but *over*-estimated in 1962–76. That the nominal GNP forecasts were on average too low in the latter period must be attributed entirely to the underprediction of the rate of increase in IPD (cf. lines 5, 6, and 10, odd columns, table 13.1).

In fact, inflation accelerated greatly in the late 1960s and, especially, in the middle and late 1970s, due to the monetary overstimulation during the Viet-

12. Weighted mean errors, in index points 1947–49 = 100, are approximately 0.8 for base levels, −1.0 for changes, and −0.1 for target levels. Note that IP is less affected by the underestimation of growth than GNP is. This is related to the fact that IP is in real units and has a smaller trend and larger cyclical component (its forecasts suffer more from turning-point errors; see Zarnowitz 1972b, pp. 194–95).

nam War and then the unusual supply and price shocks and the public and policy reactions to them. Bouts of inflation alternated with poorly managed efforts to disinflate. Real activity declined mildly in 1970, more seriously in 1974–75. The turning points in both inflation and output were generally missed; so the rises in IPD were increasingly underpredicted at the same time that the rises in RGNP were increasingly overpredicted between 1962 and 1978. These errors offset each other to produce GNP forecasts with mostly negative but on average small errors (cf. lines 2–3, 6–7, and 10–11, odd columns).

Finally, inflation peaked in 1980–81 and decreased gradually to much lower levels 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 on the whole well anticipated. RGNP growth was underestimated in 1980, when the recession turned out to be 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. However, more than half the time, in both the 1970s and the 1980s, the errors of the annual forecasts for all three variables were moderate (less than one percentage point), and on the whole the under- and overestimates balanced each other well, as can be seen from the results for 1969–89 (lines 4, 8, and 12).

13.6 Bias or Cyclical Errors?

Persistent under- or overprediction of actual values suggests a failure to avoid bias by learning from past errors. Thus it is desirable that the mean error of a set of forecasts not differ statistically from zero. Also, forecasts should be efficient, that is, uncorrelated with their own errors, else again the presence of a systematic error is inferred.

These requirements are certainly logical and indeed are often treated as almost obvious and minimal in the literature. Yet they are based on assumptions that are only too frequently shown to be false in practice, namely, that the behavioral patterns of and relations among the variables concerned are essentially time invariant and known. In reality, the processes underlying the time series to be predicted are not necessarily stable, because they reflect the changing structure and institutions of the economy and perhaps shifts in economic policies and behavior. Correspondingly, forecasting models and techniques also do not remain unchanged for long. The available samples of consistent and comparable predictions are in many cases too small to establish the existence and evaluate the importance of systematic errors. Finally, 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 and/or inefficiency even in forecasts which would be judged very good ex ante (those that come from respected professional sources or enjoy wide support when made or are based on state-of-the-art models). It seems unlikely that these forecasts are in fact systematically deficient in the sense of having persistent yet avoidable errors. More plausibly, such errors are themselves period specific and of the kind that could not be readily detected and corrected at the time when the forecasts were made.

Forecasts of inflation by individual respondents to the quarterly ASA-NBER surveys in 1968–79 provide an instructive example in this context. Tests presented in chapter 16 show a high proportion of these forecasts to be biased and inefficient, in contrast to the corresponding predictions for five other aggregative variables, where the incidence of such errors is low. In addition, there is evidence that the average forecasts of inflation from the same surveys are inconsistent with the hypothesis of unbiasedness for the period 1970–74 but not for 1975–79 and 1980–84 (Hafer and Hein 1985, esp. pp. 390–92). It makes good sense to argue that changes in inflation were particularly difficult to predict in the first half of the 1970s because of novel exogenous developments and shocks. There was the breakdown of the Bretton Woods arrangements and transition to the floating exchange regime; the imposition and elimination of wage and price controls; international food shortages and huge price increases; and the oil embargo and quadrupling of oil prices. The resulting elements of inescapable surprise must be taken into account in assessments of the generally poor forecasts of inflation rates in this period, which lagged behind the actual rates and underestimated them greatly most of the time (see chapter 14, sec. 14.5 for an analysis by subperiods).

In chapter 16, the forecasts are compared with the last data available prior to the benchmark revisions of 1976 and 1980. Keane and Runkle (1990) show that when instead the preliminary (first release) data are used, the proportions of bias and inefficiency in the one-step-ahead inflation forecasts are much smaller, so that the rationality hypothesis can no longer be rejected. Their estimation and replication work is very careful and proficient but their explanation relies too heavily on measurement errors to be really persuasive. Systematic yet unpredictable errors in inflation *data* may well have been concentrated in the first half of the 1970s like the similar large errors in inflation *forecasts*, and for much the same reasons. Early estimates have much in common with extrapolations. Otherwise, it is not clear why data errors should have been so critical for the forecasts of inflation but not for the other variables that tend to be subject to larger revisions.[13] In general, the issue of whether or

13. This is certainly true for GNP and such of its components as the change in business inventories, a series notorious for grave measurement problems and errors. Most of the data used to deflate GNP come from the components of the CPI and the producer price index (PPI), and these microdata usually do not have large revisions, except for changes in weights and seasonal factors. The revisions in the IPD series to which the ASA-NBER inflation forecasts refer are thus presum-

not forecasts are unbiased or "rational" must surely hinge on much more than the choice between preliminary and revised data.[14]

Early studies found that forecasters often underestimated changes in micro- and macroeconomic variables, and they discussed the meaning and possible sources of this phenomenon (Theil [1958] 1965, ch. 5; Zarnowitz 1967, ch. 4). But longer time series of forecasts that are now available show greater frequencies of overestimates in the recent years and little evidence of any overall bias. This is illustrated in table 13.2 for the government and private predictions of annual rates of change in nominal and real GNP and IPD, 1969–89. The mean errors of both the CEA and the ASA-NBER survey forecasts are all fractional, small relative to the corresponding standard deviations, and statistically not different from zero by conventional tests (lines 5, 11, and 16).[15] On the whole, overestimates were just slightly more numerous in the CEA set, and underestimates were slightly more numerous in the ASA-NBER set; and neither type of error was systematically larger in size (cf. lines 1–2, 6–7, and 12–13).

By far the largest errors are found in the third category, where the predicted changes differ from the actual changes in sign. Such directional or turning-point errors occurred in the annual forecasts of table 13.2 only for RGNP (line 8). They relate to some of the years of business cycle contraction and troughs, namely, 1970, 1974, and 1982, and are all positive.[16]

This suggests that it is the failure to predict business downturns that is the major shortcoming of these forecasts. More general evidence comes from comparisons of the accuracy of forecasts classified ex post by the cyclical nature of their target periods. In each case, as shown by the absolute values of their mean errors and by their standard deviations and MAEs, the forecasts for contraction and trough years have been much worse than the forecasts for expansion and peak years (cf. lines 3–4, 9–10, and 14–15). The results for quarterly predictions confirm and amplify this conclusion (chapter 18, sec. 18.2.2).

---

ably due in large measure to shifts in the GNP expenditure weights. It is easy in retrospect to relate large shifts in spending patterns to the disturbances of the 1970s. Finally, figure 1 in Keane and Runkle 1990, p. 723, suggests that the discrepancies between the initial and final estimates of IPD were greater in the first than in the second half of their sample period (1968–86), certainly in levels.

14. It cannot be taken for granted that forecasters aim to predict initial values rather than try to come closer to the true values. To the extent that revisions are systematically related to some past information, rational forecasters should be able to take them into account.

15. The ratios of the means to their standard errors are all very small (less than 0.3), but there is no good reason here to make the assumptions of independence, etc., that underlie the simple significance (t) tests.

16. For more evidence on the importance of turning-point errors, based on a larger sample of earlier annual forecasts, see chapter 14, sec. 14.3. Such errors play an even greater role in quarterly forecasts, as discussed in Zarnowitz 1967, pp. 72–80, 114–20, and Zarnowitz 1974, pp. 584–90.

Table 13.2 **Types of Error in Two Sets of Annual Forecasts of Nominal and Real GNP Growth and IPD, 1969–89**

		CEA				ASA-NBER, Median			
Line	Type of Error[a]	No. (1)	ME[b] (2)	S.D.[c] (3)	MAE[d] (4)	No. (5)	ME[b] (6)	S.D.[c] (7)	MAE[d] (8)
		Gross National Product (GNP)							
1	Underestimates	9	−0.8	0.4		13	−1.0	0.8	
2	Overestimates	12	1.0	1.1		7	1.5	1.0	
3	B.C. expansions	16	−0.0	1.1	0.9	16	−0.4	1.2	0.9
4	B.C. contractions	5	1.0	1.7	1.2	5	0.9	2.4	1.8
5	Total	21	0.2	1.3	1.0	21	−0.1	1.6	1.2
		GNP in Constant Dollars (RGNP)							
6	Underestimates	8	−0.9	0.6		9	−1.2	0.9	
7	Overestimates	9	0.8	0.6		8	0.5	0.3	
8	Directional errors	3	2.3	0.8		3	2.8	1.7	
9	B.C. expansions	16	0.1	1.0	0.8	16	−0.5	1.1	0.8
10	B.C. contractions	5	1.1	1.7	1.7	5	1.7	2.1	2.1
11	Total	21	0.3	1.3	1.0	21	0.1	1.6	1.1
		GNP Implicit Price Deflator (IPD)							
12	Underestimates	9	−1.2	1.0		10	−1.4	1.7	
13	Overestimates	11	0.9	0.7		11	1.0	0.7	
14	B.C. expansions	16	−0.0	1.0	0.8	16	0.1	1.2	1.0
15	B.C. contractions	5	−0.0	2.2	1.7	5	0.8	2.9	1.8
16	Total	21	0.0	1.4	1.0	21	−0.2	1.7	1.2

[a]Underestimates: sign P = sign A and $P < A$. Overestimates: sign P = sign A and $P > A$. Directional errors: sign $P \neq$ sign A. B.C. expansions: errors of forecasts relating to years of business cycle expansion and peaks. B.C. contractions: errors of forecasts relating to years of contraction and troughs.

[b]ME = mean error.

[c]S.D. = standard deviation.

[d]MAE = mean absolute error. Not shown where equal to the corresponding ME value without regard to sign.

13.7 Relative Accuracy

Measures of absolute accuracy, by comparing predicted and actual values, show how much the former deviate from the unattainable state of perfection (no errors). More realistic criteria are found in comparisons of the accuracy of forecasts of different types and from different sources. Some common benchmarks of predictive performance are provided by models that mechanically extrapolate information contained in the past record of the series to be predicted. The appropriate models vary with the characteristics of the variables and periods concerned. In short, forecasts are best evaluated in relative terms and by more than one yardstick.

Consider four examples of "naive models": N1, which projects forward the last observed level of the predicted variable; N2, which adds to that level the last known change; N2*, which similarly projects the average of past changes;

and N3, based on an autoregressive equation with five terms. All of eight sets in an early collection of annual forecasts of GNP and IP for periods ending in 1963 proved to be superior to N1 and N2, and all but one also to N2* and N3 (Zarnowitz 1967, table 16 and pp. 86–90; N2* in this case averages changes since 1947). A partial extension of this study through 1969 shows forecasts continuing to outperform N1 and N2*, as summed up in table 13.3, lines 1–4. However, the forecasts grew much worse relative to N2 in the middle and late 1960s. This is because extrapolations of last change in GNP are at their very best in times when no major fluctuations occur in either output or inflation. Of course, this was a transitory advantage as nominal growth is seldom so well sustained.

The relative accuracy measures in table 13.3 are ratios of RMSEs of the forecasts to the corresponding RMSEs of the selected extrapolative models. The annual CEA and ASA-NBER predictions of nominal GNP growth easily outperform the last-change projections (N2), with ratios in the range of 0.46 to 0.54 for 1962–89 and subperiods (line 5). The forecasts of real GNP growth compare still more favorably with N2 (line 8), but those of IPD inflation do worse, with ratios of 0.72 for CEA and as high as 0.98 for ASA-NBER in 1969–79 (line 11).

The ratios to N4, the projections of the moving average of changes in the last four years, are somewhat higher than the N2 ratios, but they too show the forecasts of GNP and RGNP in a strong comparative position (lines 6 and 9). For IPD, however, N4 is less demanding than N2 (line 12).

Finally, N5 is hypothetical and forward looking in that it projects the mean of actual changes in the forecast period, a statistic knowable only ex post; but it is also extremely naive in the sense that it assumes a constant prediction in each successive unit period. The RMSE of these "random-walk-with-trend" projections equals the standard deviation of the future actuals. Interestingly, N5 performs much like N4 here, being just a little weaker for GNP and RGNP and slightly stronger for IPD (cf. lines 6–7, 9–10, and 12–13).

The upshot is that the annual forecasts under study are generally much more accurate than an array of simple mechanical extrapolations. The only exception is the inflation forecasts from surveys when compared with N2 since 1964 and also with N4 and N5 in 1969–79 (lines 2 and 11–13, cols. 2 and 7). The conclusion is supported by other recent studies, notably McNees 1988b (see also table 18.1 and text below for additional results and references).

However, it can be argued that the naive models represent minimal standards. The economic models and reasoning, technical skills, professional experience, and informed judgment when combined should enable the modern forecaster to do much better. Indeed, he or she is now expected to satisfy the demand for frequent predictions of developments over sequences of several quarters into the future; the old practice of year-end forecasting for the year ahead is no longer sufficient. So quarterly multiperiod forecasts are now prepared routinely by econometric service bureaus in great detail and by many

Table 13.3 **Annual Forecasts of Nominal and Real GNP and IPD: Comparisons with Selected Naive Models, 1953–89**

Line	Ratio of RMSEs Forecast to Naive Model[a]	Average of Four Sets of Private Judgmental Forecasts (predictions of levels and changes in current dollars)			CEA (predictions of percentage change)			ASA-NBER, Median (predictions of percentage change)		
		1953–63 (1)	1964–69 (2)	1953–69 (3)	1962–74 (4)	1975–89 (5)	1962–89 (6)	1969–79 (7)	1980–89 (8)	1969–89 (9)
					Gross National Product (GNP)					
1	N1	0.41	0.28	0.33						
2	N2	0.56	1.05	0.64						
3	N2*	0.66	0.43	0.50						
4	N3	0.72								
5	N2				0.53	0.46	0.48	0.54	0.53	0.53
6	N4				0.61	0.55	0.58	0.58	0.67	0.63
7	N5				0.69	0.54	0.59	0.54	0.78	0.63
					GNP in Constant Dollars (RGNP)					
8	N2				0.47	0.34	0.41	0.46	0.42	0.45
9	N4				0.54	0.35	0.44	0.52	0.44	0.49
10	N5				0.59	0.43	0.51	0.59	0.54	0.58
					GNP Implicit Price Deflator (IPD)					
11	N2				0.72	0.72	0.72	0.98	0.74	0.92
12	N4				0.61	0.49	0.54	0.93	0.39	0.73
13	N5				0.50	0.46	0.46	0.98	0.37	0.72

Sources: Cols. 1–3: based on Zarnowitz 1967, table 16, p. 87, and Zarnowitz 1974, table 3, p. 574. Cols. 4–6; based on Moore 1983, tables 26-3 and 26-4, pp. 442–45 and 448–49; *Economic Report of the President; Budget of the U.S. Government.* Cols. 7–9: based on author's files and calculations.

[a]RMSE (root mean square error) $= \sqrt{1/n\ \Sigma\ (E_t)^2}$, where $E_t = P_t - A_t$; P_t is the predicted value and A_t is the actual value (preliminary estimate which appears in February of year $t + 1$); and Σ denotes summation over all n years covered. N1 refers to the projection of the last observed level (for year $t - 1$); N2 to that of the last observed change; N2* to that of the average change from 1947 to year $t - 1$. N3 projects the average return between the present value of the series and its past values (based on regressions of A_t on A_{t-i}, $i = 1, 2, \ldots, 5$). N4 projects the moving average of the last four observed changes (for A_{t-i}, $i = 1, \ldots, 4$). N5 assumes that the mean of the actual values in the forecast period is known and projects it each year.

[b]The forecasts come from (1) *Fortune* magazine ("Business Roundup"); (2) Harris Trust and Savings Bank; (3) Prudential Insurance Company of America; and (4) University of Missouri School of Business Administration.

business economists for an array of important macroeconomic variables. These forecasts can be tested against extrapolations from state-of-the-art times-series models, which include the univariate ARIMA (autoregressive integrated moving-average) models and the multivariate VAR (vector autoregressive) models. Chapter 18 shows that the record of such tests for three macroeconometric models and group forecasts from a business outlook survey is mixed. For some variables and periods the forecasts are less accurate than either ARIMA, VAR, or BVAR (Bayesian vector autoregressions) but the opposite is true in about two thirds of the comparisons (see table 18.3, pt. B).

Sophisticated time-series models have important lessons for forecasters on how to decompose, detrend, deseasonalize, and use the stochastic properties of the series for predictive purposes. They can help avoid bias in forecasting for processes that are reasonably stable over the periods covered. Economists' forecasts include extrapolative along with other, analytical and judgmental elements; thus comparing the errors of forecasts with the errors of corresponding projections from time-series models can yield estimates of the net predictive value of the combined nonextrapolative components of the forecast (which can be positive, zero, or negative).[17] Tests of relative accuracy based on such comparisons pose standards that may be difficult to exceed but that may not be sufficient to establish the usefulness of those forecasts that meet them. This is because the strength of the time-series models generally lies in good projections of recent trends that, however, tend to lag behind actual developments and fail to give timely signals of broad changes in the economy (turning points in growth rates and levels of income, output, prices, etc.). But it is precisely such signals that are most needed by users of short-term forecasts of general economic conditions.

13.8 Forecasting Methods and Results

During the period 1968–81, the quarterly ASA-NBER surveys regularly collected information on some methodological characteristics of the forecasts. The questionnaire asked the participants which of several listed tools they used and what the relative importance of these items was in their own work. Large majorities reported using the "informal GNP model," an eclectic and flexible approach with large elements of judgment (Butler and Kavesh 1974). This "model" actually covers a variety of procedures whereby the major expenditure components of GNP are predicted and combined into an overall forecast, in nominal and real terms. The last step usually involves various adjustments that may be iterative and are designed to make the forecast internally consistent and reasonable in light of the currently available information and beliefs. Table 13.4 tells us that over 70% of the respondents used this

17. For an early discussion and examples, see Mincer and Zarnowitz 1969 and other essays in Mincer 1969a.

Table 13.4 **Forecasting Methods Used in the ASA-NBER Quarterly Economic Outlook Surveys, 1968–70, 1974–75, and 1980–81**

	Informal GNP Model (1)	Leading Indicators (2)	Anticipations Surveys (3)	Econometric Model—Outside (4)	Econometric Model—Own (5)	Other Methods[a] (6)
			Seven Surveys 1968:4–1970:2 (496 replies)[b]			
% using	77	72	65	42	23	17
% ranking first[c]	57	14	2	7	7	8
% ranking second	13	32	24	10	7	4
% ranking lower[d]	7	24	40	26	10	5
			Six Surveys 1974:1–1975:2 (308 replies)[e]			
% using	71	49	53	56	25	14
% ranking first[c]	50	5	1	9	16	7
% ranking second	13	30	18	24	7	4
% ranking lower[d]	7	25	33	22	3	3
			Six Surveys 1980:1–1981:2 (198 replies)[f]			
% using	74	49	42	53	27	19
% ranking first[c]	56	12	1	13	13	9
% ranking second	13	21	16	14	6	5
% ranking lower[d]	4	14	25	25	7	5

Sources: American Statistical Association and National Bureau of Economic Research, *Quarterly Survey of Economic Outlook,* various issues; author's files and calculations.

[a]Write-in but often not specified.

[b]The August 1969 survey was held in connection with the ASA annual meeting and attracted a very large number of respondents (128, including 46 regular panelists). Participation in the other surveys varied from 49 to 83 and averaged 61.

[c]Most important.

[d]Ranks 3 to 6 (least important).

[e]Participation varied from 46 to 62 and averaged 51.

[f]Participation varied from 24 to 46 and averaged 31.

general approach and 50% or more ranked it as first (col. 1). These proportions remained remarkably steady while the survey participation rates declined over time (many casual forecasters who participated in the early years dropped out, leaving a much smaller core of regular forecasters only).

Leading indicators were also employed by about 70% of the survey membership in 1968–70 but later that share declined to about 50%. They were ranked second by most respondents (col. 2). Anticipations surveys received references from 65% of members in 1968–70, 42% in 1980–81, and generally lower ranks (col. 3).

Users of outside econometric models accounted for more than 40% of the early survey members and more than half of those in the 1970s and early 1980s. These forecasters preferred other methods and ranked the outside models second or lower (col. 4). About one fourth of the respondents had their own econometric models, and most of them (but perhaps surprisingly not all and not in the early years) ranked these models first (col. 5). Finally, "other methods" (e.g., time-series models) were specified by fewer than 20% of the participants and preferred by about half of them (col. 6).

The different methods tend to complement each other, for example, new readings on monthly cyclical indicators and the latest results from an investment or consumer anticipations survey may be used to modify forecasts from econometric models or the informal approach. It is therefore understandable that the predominant forecasting practice is to use various combinations of these methods or techniques in a more or less judicious fashion. Indeed, this is the single most important lesson to be drawn from the replies to the question on methods as elicited in the ASA-NBER surveys. The reported rankings differ widely, reflecting the backgrounds, interests, and preferences of the individuals; but no one method is widely treated as if it were self-sufficient and always superior to each of the others.

In an effort to establish whether the forecasters' methodological choices were associated with significant differences in predictive accuracy, I first examined regressions of the individual errors of GNP forecasts on dummy variables representing different methods, one equation for each survey and for each predictive horizon (Zarnowitz 1971, pp. 65–68). The estimates related to the early surveys with high participation rates and used alternatively the classification by first ranks only and by lower ranks as well. Few of the coefficients were found to be significant (less than one in six at the 5% level, for example). The results suggested in general an absence of systematic differences between the contributions to the forecast errors of the main listed methods.[18]

A 1975 study by Su and Su, based on the 1968:4–1973:2 ASA-NBER forecasts of absolute and percentage changes and levels of GNP, RGNP, IPD, and the unemployment rate, compared the accuracy of the respondents who ranked first the informal GNP model with those who preferred econometric models (own or outside), leading indicators, and other methods. The four groups varied in their relative performance by variable, span, and type of forecast (changes vs. levels) but none surpassed the others *consistently*.[19]

Table 13.5, based on a large number of time series of individual forecasts of rates of change in GNP, RGNP, and IPD between 1968 and 1980, presents measures of average accuracy by method that omit occasional forecasters and aggregate across predictions for the current quarter and three quarters ahead. (Providing more detail and evidence for other variables would not alter the conclusions; see Zarnowitz 1983, pp. 84–85). The differences between the average RMSEs listed in the table are, line by line, very small and of uncertain significance; indeed, when rounded off to one decimal point, all but a few of them would disappear. However, it may be worth noting that when first ranks

18. Most of the significant coefficients referred to the thinly populated and apparently inferior groups such as "other methods" and, for the first ranks, anticipations surveys.

19. See Su and Su 1975, pp. 603–5. All four subsamples generated larger errors than the consensus forecasts because of larger variances.

Table 13.5 Average RMSEs of ASA-NBER Survey Forecasts, by Methods Ranked First and Lower, 1968–80

Variable	Informal GNP Model (1)	Leading Indicators (2)	Anticipations Surveys (3)	Econometric Model—Outside (4)	Econometric Model—Own (5)	Other Methods (6)
			According to First-Ranked Method			
GNP, % change	0.96	1.00	0.99	0.89	1.09	1.15
RGNP, % change	1.14	1.24	1.22	1.05	1.25	1.27
IPD, % change	0.71	0.79	0.85	0.72	0.76	0.83
			According to Lower-Ranked Method			
GNP, % change	1.03	0.97	0.95	0.98	0.76	0.97
RGNP, % change	1.19	1.12	1.10	1.13	1.12	1.18
IPD, % change	0.76	0.72	0.71	0.71	0.72	0.87

Sources: ASA-NBER, *Quarterly Survey of Economic Outlook,* various issues; for more detail, see Zarnowitz 1983.

Note: This table covers 79 individuals in at least 12 of the 46 quarterly surveys in the period from 1968:4 through 1980:1. The entries represent averages of RMSEs of forecasts for the current survey quarter and three quarters ahead. The errors are measured as differences, percentage predicted change minus percentage actual change, for each successive nonoverlapping target quarter. Measures in lines 1–3 refer to responses of those forecasters who reported using the given method as primary (ties for the first rank are not included). Measures in lines 4–6 refer to the responses of those forecasters who reported using the given method but ranking it second through sixth.

only are considered (lines 1–3), outside econometric models tend to have the smallest errors, with the informal approach a close second. When lower ranks are used (lines 4–6), own econometric models, leading indicators, and anticipations surveys have more favorable relative positions.

This is consistent with the view that combining the different procedures helps, particularly when done by experienced forecasters. Thus our sample measures indicate that subscribers perform somewhat better than model proprietors on average over time, and the probable reason is this: the former group is dominated by large companies using well-known econometric service bureaus and their own professional staffs, whereas the latter group includes some individuals who are exclusive users of their own models and some teams of experts selling their large-model forecasts and advice.

Some broadly corroborative evidence is also available from other sources. According to the annual surveys of the National Association of Business Economists (NABE) in 1975–79, 52%–60% of members preferred "eclectic judgmental" methods, and 22%–28% preferred "eclectic econometric" methods (Conlan and Wickersham 1982). A special mail survey sent to the Blue Chip forecasters in 1987 showed the following mix of average contributions to predictions of real growth, inflation, and interest rates: judgment, 48%; econometric model, 28%; time-series analysis, 24% (based on more detailed figures in Batchelor and Dua 1990, p. 5). Even the organizations with their own large-scale econometric models assigned sizable weights to judgment (20%–50%, on average about 30%) and other elements such as time-series methods, current data analysis, and interaction with others (10%–20%). Thus, these forecasters estimated the contributions of their models as such at 45%–80% (on average 60%).[20]

Some of the lists mix techniques and theories. For example, the NABE members' classification in 1980 and 1981 includes, in addition to the large judgmental and econometric groups (averaging 49% and 26%), "rational expectations" (12%), "monetarists" (6%), and other or nonrespondents. Batchelor and Dua (1990) report on an attempt to cross-classify the Blue Chip forecasters by "ideology" (Keynesian, monetarist, supply-sider, RE, Austrian, other) and "technique" (three categories, as noted above). They find some support for the inference that the Keynesian-econometric combination had an advantage over others, but note that individuals in their sample generally relied on more than one technique and used elements of more than one theory. Also, Keynesian models and econometric methods were developed earlier than the modern versions of the other theories and methods, so they may have gained adherents with more practical and diverse experience.[21]

13.9 Search for the Best and the Complementarity of Suboptimal Forecasts

The classical research strategy of economists looking into the future is to form conditional expectations based on an "optimal" model. This involves the use of the preferred theory of the behavior of economic agents as constrained by the available resources and the institutional framework; identification of the endogenous variables and the relationships among them; specific assumptions about economic policies and exogenous events or developments; approximations with existing data, statistical estimation of the model parameters, and derivation of predictions. This line of attack led to the macroeconometric models and forecasts.

But macroeconomic theories differ and no one is demonstrably superior and generally accepted. The complexity and changing dynamics and structure of the economy make it costly and difficult to collect the required information and learn from it on a current basis. Testing of the theories is impeded and ideological differences persist. Even substantial agreement on fundamentals

20. See McNees 1981, p. 7. The weights come from 11 sources of macroeconomic forecasts: BEA, Chase, DRI, General Electric Co., Georgia State University, Kent Economic Institute, Manufacturers Hanover Trust, RSQE (Michigan), Townsend-Greenspan & Co., UCLA, and Wharton.

21. The weights placed on the listed theories by the average responses to the May 1987 Blue Chip survey were as follows: Keynesian, 43%; monetarist, 20%; supply side, 12%; rational expectations, 8%; Austrian 4%; other, 13%. Batchelor and Dua examined annual forecasts of real growth, inflation, and the Treasury bill rate made by 44 respondents on selected dates in 1976–86 (1990, pp. 4–10).

of rational behavior is not sufficient to resolve conflicts of views on what constitutes credible restrictions in the econometric models of the economy. The selections of endogenous versus exogenous variables are particularly controversial.

Moreover, some of the best-known macroeconometric models are so very large that they are unwieldy and difficult to comprehend, often posing excessive data requirements and resorting to ad hoc theories and arbitrary assumptions in dealing with detailed relations about which little is known. Many features of the existing models viewed as "Keynesian" are not acceptable to critics of diverse persuasion: monetarist, rational expectations, public choice, supply side. This applies notably to the treatment of economic policies as exogenous, basically benevolent, and effective not only in principle but also often in practice. But the critics have yet to produce their own, and evidently better, econometric models of the economy.

The interest of academic economists in practical econometric forecasting, never strong to begin with, was much reduced by the recent controversies, which partly explains the rise in popularity within the profession of new statistical methods of univariate and multivariate time-series prediction. Econometric forecasts were compared successively with simple autoregressive (AR), ARIMA, VAR, and BVAR forecasts. The challengers claimed that the time-series models have the advantage of low costs and replicability but yet are competitive with the best-known complex and expensive econometric models with respect to many, though not all, variables, horizons, and periods; or that the econometric forecasts are inefficient in that lower errors can be obtained by combining them with some time-series models (Nelson 1972; Cooper 1972; Cooper and Nelson 1975; Lupoletti and Webb 1986; Litterman 1986).

In their countercriticism, econometricians noted correctly that only their forecasts have the potential advantages of being based on models with identifiable structures and specific assumptions about exogenous variables and the possibility to explain and simulate as well as predict. However, they also argued that their models require the use of prior knowledge in structural specification and inspection of the equation residuals before each forecast is made. The charge against the time-series models is that they fail to take proper advantage of economic theory, may be restricted to too few variables and too many lags, and are unlikely to predict well over longer horizons (Howrey, Klein, and McCarthy 1976; Runkle 1987). A lively debate about the methods of evaluation, the linkages between, and the relative performance of time-series and econometric models has continued for years and shows no signs of exhaustion.[22]

22. The following papers are cited for their innovative nature or because they review the subject (some are accompanied by several comments): Zellner and Palm 1974; Wallis 1977, 1989; Armstrong 1978; Zellner 1979; Fildes 1985; Longbottom and Holly 1985; McNees 1986; Dhrymes and Peristiani 1986; Clemen and Guerard 1989.

At the same time that the econometricians and time-series analysts engaged in a competition guided by the principle of constructing optimal predictive models, considerable work was being done on combining multiple individual forecasts of various types. This research demonstrates that such combinations generally improve forecast accuracy, often substantially and by very accessible and inexpensive methods, including simple averaging (see Clemen 1989 for a survey with annotated bibliography). I first presented and discussed the evidence on gains from aggregating individual GNP forecasts in 1967 (pp. 5 and 123–26); a more recent and more comprehensive analysis is given in chapter 15 of this volume. Much has been learned from two decades of effort to develop a theory of optimal forecast combinations, which however does not promise a single best rule but rather suggests different procedures depending on the underlying assumptions and purposes (Winkler 1989).[23]

The idea of combining forecasts is to some critics inconsistent with the principle of optimal information-processing and modeling: an econometric structure that does not "emcompass" what can be predicted by a time-series extrapolation, for example, is simply misspecified (Chong and Hendry 1986). Further, combining (like time-series models) can result only in unconditional forecasts and may generate internal inconsistencies, for example, predictions of GNP components that do not add up to predictions of total GNP.

In a world in which economic processes and relations tended to be stable and identifiable from good data promptly available at low cost, pooling of information would always be preferred to pooling forecasts (which in this case should not vary much anyway). But ours is a very different world where "economic change is a law of life" (Burns 1968, p. 226); new surprises and uncertainties continually arise, and valuable information is costly and at no time exclusively possessed by any single expert or embodied in any single model. Timely short-term forecasts for the economy under such conditions can hardly afford the costs of collecting all the relevant data and knowledge. Thus combining forecasts may be justified here as a practical way to aggregate the pieces of information that are available to forecasters, and the procedure can be formalized along Bayesian lines (Winkler 1989).

In particular, combining bona fide outside forecasts with different characteristics is an expedient method for a decision maker to reduce the large-error risk associated with relying on one particular model or one individual's judgment.[24] Here then is an appropriate role for *users* (and collectors and analysts) of the forecasts. However, the essential function of *makers* of forecasts is very different, namely, to *add* some predictive value to the sum of diverse infor-

23. The literature advanced from combinations of unbiased one-step forecasts with weights constrained to sum to 1 (Bates and Granger 1969) through unconstrained least squares (Granger and Ramanathan 1984) to Bayesian prior-information and shrinkage techniques (Clemen and Winkler 1986; Diebold and Pauly 1990). For a study of particular interest to macroeconomists, see Bischoff 1989.

24. A close and often noted analogy is with an investor's strategy to reduce risk through portfolio diversification.

mational inputs acquired from outside (data, tools, interactions with others). Unless a forecaster produces some such "value-added," his or her product will not be sufficiently differentiated to make a contribution to the combined forecast and to be of continuing interest to informed users.

13.10 Model and Judgment

At the most basic level, two ingredients can be distinguished that are blended in the making of almost any macroeconomic forecast: (1) some more or less systematic technique or model and (2) judgment in choosing and modifying ingredient (1) and adjusting its results. Some forecasters wish to reduce judgment to the choice of the procedure and rely mainly on the model in the interest of objectivity, replicability, and avoidance of biases of perception and assessment. Uses of time-series models that require little or no individual fine-tuning, such as unrestricted univariate or vector autoregressions, provide good examples. Others exercise their judgment much more extensively so as to apply prior knowledge and experience in diagnosing the changing conditions and flexibly adapting the current forecast to them. This is illustrated by the practices not only of many business economists who have no formal models of their own but also of those econometricians who often judgmentally adjust many predictions generated by their models in attempts to improve their accuracy.

The major role of such constant-term adjustments in macroeconometric forecasting of the 1960s is amply demonstrated in studies by Evans, Haitovsky, and Treyz (1972) and Haitovsky and Treyz (1972) (for an interpretation, see also Zarnowitz 1972b). First, the ex ante predictions by teams equipped with the then-representative large models, Wharton and Office of Business Economics (OBE) (Department of Commerce), are much more accurate with than without judgmental adjustments (i.e., in their final form as issued, XA*, rather than in the intermediate, unpublished stage as generated mechanically from the models, XA). This is summed up in table 13.6, which also indicates that the reductions in the MAEs produced by the adjustments are particularly large for the shortest predictive spans (lines 1–4, cols. 1–3).[25]

Second, the errors of ex ante forecasts are on average smaller in absolute size than the errors of the corresponding ex post forecasts that incorporate the same adjustments: the MAE ratios XA*/XP* tend to fall in the 0.6–0.8 range (lines 5–8, cols. 1–3). This is surprising, since XP* use the reported realized values of the exogenous variables and should on this account be more accurate than XA*, which use the projected values of these variables. Although the forecasters' adjustments are themselves a source of errors that may either reinforce or offset the errors in the models and external extrapolations, their net

25. The results for some other variables are similar. See, e.g., Zarnowitz 1972b, table 6 and pp. 218–22.

Table 13.6 **The Effects of Judgmental Adjustments on the Accuracy of Forecasts with Several Macroeconometric Models**

		Two Models, 1967:2–1969:3[b]			Four Models, 1980s			
Line	Variable[a]	1Q (1)	4Q (2)	All (3)	1Q (4)	4Q (5)	8Q (6)	All[c] (7)
		MAE Ratios: XA/XA*[d]			*% of Predictions Improved by Judgment*[e]			
1	GNP	0.2	0.6	0.4	66*	47	65	59
2	RGNP	0.2	0.9	0.6	55	50	50	52
3	UR	0.3	0.6	0.5	60	48	50	53
4	Total	0.2	0.7	0.5	62*	57	58	59
		MAE Ratios: XA/XP**[f]			*% of RMSEs Reduced by Judgment*[g]			
5	GNP	0.6	1.1	0.8	100	50	50	67
6	RGNP	0.7	0.7	0.8	75	25	25	42
7	UR	0.8	0.8	0.7	75	50	50	58
8	Total	0.7	0.9	0.8	76	68	63	69

Source: Cols. 1–3; based on Haitovsky and Treyz 1972, table 1 and p. 319. Cols. 4–7: based on McNees 1990, table 4 and pp. 46–48.

[a]GNP = nominal GNP; RGNP = real GNP; UR = civilian unemployment rate. Total refers to averages for the same 3 variables (cols. 1–3) and for 21 variables covered (cols. 4–7).

[b]Averages for the Wharton model and the OBE model forecasts. 1Q and 4Q denote one quarter ahead and four quarters ahead, respectively. "All" refers to averages for 1Q, 2Q, 3Q, 4Q, and 1-year-ahead forecasts.

[c]Refers to averages for 1Q, 4Q, and 8Q (eight quarters ahead) forecasts.

[d]MAE = mean absolute error. XA* = judgmentally adjusted ex ante forecasts, XA = unadjusted ex ante forecasts.

[e]Percentage of times that adjusted predictions were more accurate than those generated mechanically. Total number of observations for 1Q forecasts in line 4 is 841. An asterisk after a number indicates that it is significantly different from 50 at the 90% confidence level.

[f]XP* = judgmentally adjusted ex post forecasts. XA* and XP* incorporate the same adjustments.

[g]RMSE = root mean square error. The number of the RMSEs in each of cols. 4–6 is 4 (lines 5–7) and 71 (line 8).

effect was apparently to partially compensate for the other inaccuracies. Both outside information and judgment can help correct for errors that an unaided model would commit, but uncontrolled interactions between the different categories of errors may present a serious problem.[26]

Comparisons of Wharton and OBE forecasts of GNP and RGNP in 1966–69 with largely judgmental forecasts by business economists yield mixed re-

26. On cases where the ex post forecast errors exceed the ex ante ones without adjustments, which suggests that model misspecifications are more than offset by errors in the exogenous inputs, see Zarnowitz 1972b, pp. 27–28 (also, cf. comments by Okun and Eckstein in Zarnowitz 1972b, pp. 319–22).

sults (Zarnowitz 1972b, tables 7 and 8, pp. 222–27). The adjusted ex ante forecasts of Wharton and OBE hold an edge over two sets of the other forecasts but not two others (including ASA-NBER group median predictions). The ex post forecasts are somewhat less accurate.

These comparisons are of limited value because the available samples of forecasts that can be matched are small, and they refer to old models that have been much revised since. Some new results are presented in McNees 1990 for "four prominent macroeconometric forecasters . . . who . . . have provided data on both their publicized (adjusted) and mechanical (unadjusted) forecasts (pp. 46–47). Table 13.6 gives a summary (cols. 4–7). Judgment improved 55%–66% of the individual predictions for one quarter ahead and reduced 75%–100% of their RMSEs. Here the adjustments tend to receive much help from data on weekly and monthly indicators. For longer spans, the proportions are considerably lower but in general still 50% or higher overall.[27]

Counts of how often judgmental adjustments improved accuracy do not tell us how large the reductions in the averages of absolute or squared errors were. Even so, the new measures seem to leave judgment a smaller role than the old ones do. This could be due to the more recent models being better specified or including more efficient predetermined rules for adjusting residuals or some other reasons. An analysis of the relation between errors of published (adjusted) forecasts and errors of mechanical forecasts suggests that the judgmental adjustments, although mostly helpful, are more often than not too large; the forecasters would do better if they relied on them somewhat less and on their own models somewhat more (McNees 1990, pp. 49–51). But additional evidence is needed to clarify this important aspect of the actual use of macroeconometric models in forecasting.

In my view, it is still largely valid to conclude, as past research did, that the contributions of professional judgment and experience to the accuracy of macroeconomic predictions tend to be both important and positive. After all, an economist's knowledge and analysis of current developments, which a model cannot have, should be able to improve on the mechanical forecasts from that model. This need not at all be inconsistent with psychologists' findings that cast doubt on the value of untrained and unmotivated "common sense" in experimental predictive environments.

13.11 New Comprehensive Comparisons

A detailed study of the forecasting performance of the NBER-ASA Quarterly Economic Outlook Survey 1968:4–1990:1 has been completed very recently (Zarnowitz and Braun 1991). The results confirm that the dispersion across the individual participants' forecasts is typically large and rising with

27. Compare col. 4 with cols. 5–7 in the table. A conspicuous exception is the RMSEs for RGNP in line 6, cols. 5–7. For more detail on more variables, see McNees 1990, table 4.

the length of the predictive horizon. Errors of the average change forecasts cumulate over longer spans with great regularity for a variety of time series. Errors of marginal change and level forecasts, too, often increase with the distance to the target quarter, although by smaller margins and less regularly.

The more autocorrelated variables such as real GNP and consumption are much easier to forecast, and are much better predicted, than variables with high random variability such as residential investment and change in business inventories (all forecasts of series with trends refer to percentage changes). Inflation was underestimated and poorly predicted by most forecasters most of the time.

Simple averaging across the corresponding responses to each successive survey results in group mean forecasts that are generally much more accurate than the majority of individual forecasts. However, for some variables and periods the combinations work much better than for others. The more differentiated and the more complementary the information embodied in their components, the better are the group mean (consensus) forecasts.

Table 13.7, which covers rates of change in GNP, RGNP, and IPD from 1968:4 to 1990:1, compares the mean individual and the consensus forecasts from the NBER-ASA surveys with some representative econometric and time-series forecasts. The econometric predictions are those based on the model of the University of Michigan Research Seminar in Quantitative Economics (RSQE), the longest available series of consistent forecasts of this type. The BVAR forecasts use the five-variable, six-lag quarterly model introduced in chapter 12 (with M2). The Sims probabilistic model is also of the BVAR type

Table 13.7 Nine Sets of Forecasts Ranked according to Their Average RMSEs, Three Variables, 1968:4–1990:1

		Gross National Product (GNP)		GNP in Constant Dollars (RGNP)		Implicit Price Deflator (IPD)	
Line	Forecast	ARMSE (1)	Rank (2)	ARMSE (3)	Rank (4)	ARMSE (5)	Rank (6)
1	NBER-ASA median	1.90	4	1.94	7	1.53	7
2	NBER-ASA consensus	1.586	1	1.58	3	1.21	5
3	Michigan (RSQE)	1.98	5	1.87	5	1.42	6
4	BVAR(A)	2.69	8	1.90	6	1.62	8
5	BVAR(B)	1.89	3	1.40	1	1.03	3
6	Sims(A)	2.30	7	2.08	8	.97	2
7	Sims(B)	1.594	2	1.47	2	.66	1
8	Sims-Todd ARIMA(A)	3.05	9	2.26	9	1.69	9
9	Sims-Todd ARIMA(B)	2.03	6	1.60	4	1.09	4

Source: Zarnowitz and Braun 1991, table 29.

Note: ARMSE (average root-mean-square error) is computed by taking the mean of the RMSEs across the five spans 0–1, . . . , 0–5. The smallest ARMSE is ranked 1, and the largest ARMSE is ranked 9, for each of the three variables.

but allows time variation in coefficients and forecast error variance; it is a nonlinear, nine-variable, five-lag model (Sims 1989; for an earlier version, see Litterman 1986). The univariate ARIMAs are as specified in Sims and Todd 1991. All these time-series models are estimated with the presently available data that incorporate all revisions; hence the forecasts based on them are in this sense ex post. But the forecasts are generated sequentially, using only the information preceding the date of the forecast.

Two alternative assumptions, A and B, are employed for the comparisons in table 13.7. Variant A is that the last-known values of the variables concerned refer to the previous quarter, $t-1$; variant B is that they refer to the current quarter, t. A is preferred because the quarterly data for t are not known to the real-time forecasters, but B is to some extent justified because the forecasters do know and use some monthly and weekly data released in quarter t (and the latest economic news generally). The truth falls somewhere in between but probably more often closer to A than B, for two reasons: (1) the forecasters' information about the most recent developments is limited and deficient; (2) the forecasters use preliminary data, and the time-series models use revised data.

For variant A comparisons, the average RMSEs of the consensus (group mean) survey forecasts are the lowest for GNP and RGNP and the second lowest for IPD, following the Sims (A) model (lines 1–4, 6, and 8). The variant B comparisons are rather strongly biased in favor of the ex post forecasts with time-series models. The ARMSEs are all much lower for the variant B predictions than for their variant A counterparts (cf. lines 4, 6, and 8 with lines 5, 7, and 9). When all nine sets of forecasts listed in table 13.7 are considered, the Sims (B) model ranks 2, 2, and 1 for GNP, RGNP, and IPD, respectively. The corresponding ranks of BVAR (B) are also high: 3, 1, and 3. The ARIMA forecasts tend to be less accurate.[28]

13.12 A Preview

The last part of this book develops several themes already introduced and some additional ones. Chapter 14 argues that the accuracy and properties of forecasts depend heavily on the economic characteristics of the periods covered but only weakly and not systematically on the differences among the forecasters. Offsets between errors in the corresponding predictions of real growth and inflation are demonstrated and analyzed. Multiperiod quarterly forecasting is shown to pose much greater difficulties than annual forecasting.

Chapter 15 discusses the variety of predictions covered by quarterly business outlook surveys. Combining individual forecasts from different professional sources—business analysts, academic economists, corporate execu-

28. Note that these results conceal the differences between the forecast horizons, which are sometimes important, and that the rankings for some other variables differ considerably. Thus, the Michigan forecasts rank higher for the longer spans and are best for the rate of unemployment.

tives—can result in significant gains. Thus the group mean forecasts are on the average over time more accurate than most of the component sets. But there is also a moderate degree of consistency in the relative performance of a large number of the survey members.

Chapter 16 presents extensive results from testing for bias and serially correlated errors in a collection of time series of quarterly forecasts with several horizons and for several variables. It argues against the presumption of rationality in the sense that one should not expect the macroeconomic forecasts to be typically either uniform, unbiased, or self-fulfilling. The tests are more favorable to composite group forecasts than to most of the individual forecast sets, and less favorable to predictions of inflation than to those of other variables, including RGNP growth and unemployment.

Chapter 17 uses unique survey data on matched point and probabilistic forecasts of inflation and GNP growth to study how the degree of consensus among forecasters is related to the degree of uncertainty as revealed by the diffuseness of the appropriate probability forecasts. This means that the disagreement among forecasters tends to understate uncertainty but that rising disagreement often indicates rising uncertainty. Also, there is evidence that expectations of higher inflation generate greater uncertainty about inflation, and that the latter has adverse effects on real growth.

Finally, chapter 18 finds no evidence that U.S. macroeconomic forecasts have grown systematically worse, that is, less accurate, more biased, or both (as some critics have charged). True, large errors in predictions of both real growth and inflation occurred in some recent years (the mid- and late 1970s and early 1980s) but these were times of high concentration of unanticipated shocks and setbacks. The major failures of forecasting are shown to be related mainly to the incidence of slowdowns and contractions in general economic activity. Accordingly, progress in forecasting will require better handling of the difficult problem of turning-point prediction. There is need to combine econometric and time-series models with uses of leading indicators to reduce the length and variability of the lags in recognizing recessions (see also Zarnowitz and Moore 1991).