Volume Title: Business Cycles, Indicators and Forecasting

Volume Author/Editor: James H. Stock and Mark W. Watson, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-77488-0

Volume URL: http://www.nber.org/books/stoc93-1

Conference Date: May 3-4, 1991

Publication Date: January 1993

Chapter Title: Twenty-two Years of the NBER-ASA Quarterly Economic Outlook Surveys: Aspects and Comparisons of Forecasting Performance

Chapter Author: Victor Zarnowitz, Phillip Braun

Chapter URL: http://www.nber.org/chapters/c7189

Chapter pages in book: (p. 11 - 94)

# 1 Twenty-two Years of the NBER-ASA Quarterly Economic Outlook Surveys: Aspects and Comparisons of Forecasting Performance

Victor Zarnowitz and Phillip Braun

Human action has to a large extent always been oriented toward the future. Since ancient times, men and women hoped to outwit fate and survive by magic divination; they also hoped to outwit nature and others by shrewd calculation. Attempts to predict the future, therefore, are as old as magic, but they are also as old as commerce, saving, and investment. Their motivation must always have been largely economic, despite the inevitable frustrations of economic forecasting.

Great foresight in business matters is presumably highly profitable and rare. Its possessor will do well to exploit it directly for personal enrichment and hence should not be inclined to offer its products to the public in the open market. An economist who perceives competitive markets as working with reasonable efficiency should not expect any forecasts of stock prices or interest rates to be both freely traded and consistently much better than average. Forecasting macroeconomic aggregates such as real GNP and its major expenditure components is likely to have less potential for direct profitability than forecasting financial variables. Hence, it is presumably less vulnerable to that old American adage rebuking expert advisers: "If you're so smart, why ain't you rich?" (cf. McCloskey 1988).

For reasons explained in section 1.1 below, to be interesting and robust, macroforecast assessments should cover a broad range of forecasters, variables, and economic conditions. The forecasts must be explicit, verifiable,

Victor Zarnowitz is professor emeritus of economics and finance at the University of Chicago, Graduate School of Business, and a research associate of the National Bureau of Economic Research. Phillip Braun is assistant professor of finance at Northwestern University, Kellogg Graduate School of Management.

The authors thank Ray Fair, Christopher Sims, Allen Sinai, James Stock, Mark Watson, and the participants in the discussion at the NBER meeting on 4 May 1991 for helpful comments; Christopher Culp for research assistance; Cynthia Davis for typing; and the NBER for financial assistance.

and sufficient to allow a responsible appraisal. Unfortunately, most of the available time series of forecasts are short, and none are free of some gaps, discontinuities, and inconsistencies. Relying on a small sample of specific forecasts from an individual source risks overexposure to isolated hits or misses due to chance. It is therefore necessary to concentrate on a set of forecasts from numerous and various sources. This is likely to improve the coverage by types of information and methods used as well.

The way to collect the required data is to conduct regularly, for a sufficiently long time and with appropriate frequency, a survey that would be reasonably representative of the professional activities of macroeconomic forecasters. A joint project of the National Bureau of Economic Research (NBER) and the Business and Economic Statistics Section of the American Statistical Association (ASA) had the purpose of accomplishing just that. The NBER-ASA survey assembled a large amount of information on the record of forecasting annual and quarterly changes in the U.S. economy during the period 1968:IV–1990:I (eighty-six consecutive quarters). It reached a broadly based and diversified group of people regularly engaged in the analysis of current and prospective business conditions. Most of the economists who responded came from corporate business and finance, but academic institutions, government, consulting firms, trade associations, and labor unions were also represented. The forecasts covered a broad range of principal aggregative time series relating to income, production, consumption, investment, profits, government purchases, unemployment, the price level, and interest rates. The surveys also collected data on the methods and assumptions used by the participants and on the probabilities that they attached to alternative prospects concerning changes in nominal or real GNP and the implicit price deflator.

The NBER-ASA data have their shortcomings, most important of which are probably the high turnover of participants and the large frequency of gaps in their responses. The data collected represent a mixture of public and private predictions. The survey members, generally professional forecasters, were identified by code only. Their anonymity helped raise the survey response rates but may have had otherwise ambiguous consequences (encouraging independence of judgment or reducing the sense of individual responsibility?).

The initiative to develop and maintain the quarterly NBER-ASA survey was strongly motivated by the desire to make it "the vehicle for a scientific record of economic forecasts" (Moore 1969, 20). The expectation that such a survey would be of considerable service to both the profession and the public was shared by Moore with others who helped implement his proposal (including one of the authors of this paper, who had the responsibility for reporting on the NBER-ASA survey during the entire period of its existence). In retrospect, it seems fair to say that the assembled data do indeed provide us with rich and in part unique information, which can help support much-needed research on the potential and limitations of forecasting economic change.

Twenty-two years of a survey that attracted numerous responses from a variety of sources each quarter add up to a mass of information about the processes and results of macroeconomic forecasting. Although many studies have already used some of this material, much of it remains to be explored. This report is the first to examine all the variables included in the NBER-ASA forecasts, for all horizons and over the entire period covered. It concentrates on the properties of the distributions of summary measures of error, by variable and span of forecast, viewed against the background of descriptive statistics for the predicted time series. Other subjects of interest include the role of characteristics and revisions of "actual" data in the evaluation of the forecasts, differences by subperiod (roughly the 1970s vs. the 1980s), the relation between the individual and the group mean or "consensus" forecasts from the surveys, the comparative accuracy of the survey results and predictions with a well-known macroeconometric model, and comparisons with forecasts from state-of-the-art multivariate and univariate time-series models.

Section 1.1 of this paper examines some general problems and the history of forecast evaluations and surveys. Section 1.2 presents the NBER-ASA data and the methods used. Sections 1.3–1.5 discuss the results of the analysis and form the core of the paper. Section 1.6 draws the conclusions.

## 1.1 The Diversity of Forecasts and Their Evaluation

### 1.1.1 Some Reflections on Predictability and Uncertainty

It can be readily observed that, at any time, predictions of a given variable or event can and in general do differ significantly across forecasters. Indeed, modern macroeconomic forecasts display a great diversity, which must be taken into account in thinking about how to assemble and evaluate the related data.

Although changes in the economy are predicted primarily to meet the demand for forecasts by public and private decision makers, they are also predicted to test theories and analytic methods and to argue for or against points of policy. Some conditions and aspects of the economy are much more amenable to prediction than others. Furthermore, individual forecasters differ with respect to skills, training, experience, and the espoused theories and ideologies. They compete by trying to improve and differentiate their models, methods, and products. They respond to new developments in the economy and new ways to observe and analyze them. In sum, there are both general and specific reasons for the observed diversity of forecasts.

Comparisons among forecasts that are differentiated in several respects are difficult yet unavoidable. The quality of a forecast is inherently a relative concept. Common standards of predictive performance must therefore be applied to properly classified forecasts along each of the relevant dimensions.

Surely, the main value of a forecast lies in its ability to reduce the uncertainty about the future faced by the user. In general, a forecast will perform better in this regard the smaller and closer to randomness its errors are. However, the value of a forecast depends not only on its accuracy and unbiased nature but also on the predictability of the variable or event concerned. Some events and configurations of values are common, others rare. Where the probability of occurrence for the forecasting target is high, uncertainty is low, and prediction is easy but not very informative. Where that probability is low, uncertainty is high, and prediction is difficult but potentially very valuable (cf. Theil 1967).

For example, total stocks of the nation's wealth and productive capital normally change little from one month or quarter to the next, barring a catastrophic war or a natural disaster, and so can be predicted with small relative errors. Much the same applies to other typically "slow" stock variables such as total inventories of goods or monetary aggregates and the overall price level (but not in periods of rapid inflation!). In contrast, income and expenditure aggregates represent "fast" flow variables, some of which (e.g., corporate profits, investment in plant and equipment, housing starts, and change in business inventories) are highly volatile over short horizons and apt to be very difficult to forecast accurately. Rates of change in indexes of price levels fall in the same category.

There are also situations that are unique or nearly so where no objective or subjective probabilities based on past history or experience are believed to apply and where "true" (nonergodic) uncertainty rules (as in Knight 1921, 233). According to Keynes (1936, 149), "Our knowledge of the factors which will govern the yield of an investment some years hence is usually very slight and often negligible," yet businesspeople must make decisions to make or buy plant and equipment despite this recognized state of ignorance. In economics, as in history, statistical-stochastic methods have limited applicability (cf. Hicks 1979; Solow 1985). Forecasters cannot afford to be deterred by such considerations and assume some predictability throughout, never full uncertainty.

Across many variables, uncertainty depends on the "state of nature" (more explicitly, on the state of the economy or the phase of the business cycle). Thus, it is much easier to predict continued moderate growth once it is clear that the economy has entered a period of sustained expansion than it is to predict the occurrence and timing of a general downturn after the expansion has lasted for some time and may be slowing down.

Influential public macro forecasts could in principle be either self-invalidating or self-validating. Thus, if the government believes a forecast of a recession next year, it might succeed in stimulating the economy so as to make the expansion continue. On the other hand, if consumers generally come to expect a recession because of such a forecast, individuals may try to protect themselves by spending less now and dissaving later when the bad times arrive. Businesspeople, acting on similar expectations, may reduce investment expenditures and financing, production, and inventory costs. But such actions, although individually rational, would collectively help bring about the recession no one wants.

Indeed, an early theoretical monograph on forecasts of general business conditions concluded, on these grounds, that they *cannot* be accurate, particularly if they are made public (Morgenstern 1928). However, it is not necessarily true that a known forecast must be falsified by agents' reaction to it, even if that reaction does affect the course of events. Conceptually, the reaction can be known and taken into account for bounded variables related by continuous functions (Grunberg and Modigliani 1954).[1] But the public prediction can be correct only if the corresponding private prediction is correct, which of course is often not the case. Forecasting remains difficult whether or not its results are published. The premise of a generally shared belief and confidence in a commonly held forecast is so unrealistic as to deprive theoretical exercises based on it of much practical interest.

### 1.1.2 A Brief History of Forecast Appraisals and Surveys

Qualitative judgments about contemporary levels of, and changes in, general business activity are among the oldest economic data. A compilation of such records provided partial evidence for the NBER work on identifying and dating the business cycles of history (Thorp 1926; Burns and Mitchell 1946). A look at these "business annals," which go back to the 1830s, reminds one of the importance of public perceptions and expectations concerning aspects of general economic and financial activity: employment, production, prices, interest rates.

This expectational element in the dynamics of economic life has probably long attracted great attention from students of current events and men and women of affairs. It did not much concern those early theorists, who were preoccupied with problems of long-run static equilibrium. But some prominent economists in the classical tradition stressed the role that variations in expectations and "confidence" play in business cycles (Marshall), or hypothesized the occurrence of sequences of overoptimism and overpessimism (Pigou), or attributed to bankers and entrepreneurs predictive errors resulting in malinvestments (Hayek). Keynes and some of his later followers elaborated on the destabilizing role of uncertainty. Along with the formal models of interacting economic processes came the theories of expectation formation, first that of adaptive and later that of rational expectations. In the last twenty years or so, incomplete information and expectational errors acquired prime importance in many models of economists of various persuasions (monetarist, new-classical, new-Keynesian). The corresponding literature grew rapidly.

Lack of quantitative data has long hampered the progress of economics,

1. Interestingly, Morgenstern's monograph and Grunberg and Modigliani's paper are in a sense precursors of the contemporary rational expectations models in which behavior follows forecasts that are consistent with the assumptions of the models and free of any systematic errors.

causing empirical work and tests to lag well behind the formulation of theories and hypotheses. Numerical data on forecasts and expectations are particularly scarce, except for the very recent period of great expansion in economic and financial prediction and consulting activities. Hence, the literature on macroeconomic forecasting has a brief history, although it too has grown rapidly of late.[2]

The first forecasting services in the United States to gain considerable success date back to the years immediately preceding World War I and the 1920s. They used lead-lag relations to predict business-cycle turning points, relying mainly on the tendency of stock prices to lead and short-term interest rates to lag business activity. The sequence, best known as the Harvard "ABC" curves, had a basis in theory and fact, but it was a crudely oversimplified predecessor of the indicator system subsequently developed at the NBER. It performed rather well in the period 1903–14 and in the depression of 1920–21, and it would have applied generally in recent times as well (cf. Moore 1969), but the Harvard service failed to foresee the onset and extent of the Great Depression, which doomed this and related forecasting efforts. A 1988 postmortem study, using the Harvard data and modern vector autoregressive (VAR) model techniques, concludes that the large declines in output that followed the 1929 stock market crash could not have been forecast (Dominguez, Fair, and Shapiro 1988).[3] This, however, is disputed by a very recent paper that applies the Neftci sequential analysis method to the Harvard index (Niemira and Klein 1991).

Monthly forecasts from six sources for the period 1918–28 were scored for accuracy in Cox (1929), to our knowledge the first methodical appraisal of ex ante predictions of U.S. business activity. Cox found evidence of a moderate forecasting success despite the poor showing at the 1923–24 recession.

The earliest compilation of quantitative macro forecasts, so far as we can tell, was the informal survey conducted since 1947 by Joseph Livingston, the late syndicated financial columnist based in Philadelphia. Twice a year he collected predictions of such variables as industrial production and the consumer price index and summarized the results in a business outlook column published in June and December. The forecasters were mostly business and financial economists, but some academics were also included. The Livingston data represent a unique and valuable source of information on forecasts for the early post–World War II period, and, in the 1970s, they began to be widely used in research, primarily on price expectations. But Livingston adjusted his published "consensus forecasts" (means of the collected individual predictions) in an attempt to take into account any large revisions in the actual data that may have occurred between the mailing of his questionnaire and the sub-

2. The same applies to the literature on microeconomic prediction, which is additionally restricted by the fact that much of the material on micro forecasts is confidential.

3. Forecasts of a Yale service developed by Irving Fisher were not better in 1929 than those of the Harvard service, developed by Warren Persons (see Dominguez, Fair, and Shapiro 1988).

mission of his column to the press. Carlson (1977) recalculated the semiannual Livingston forecasts of CPI and WPI inflation rates for 1947–75 from the original data so as properly to reflect the timing of the predictions and the information incorporated in them.[4]

As quantitative macroeconomic data and forecasts began to accumulate in the 1950s and 1960s, valid examinations of the accuracy and properties of the latter became increasingly possible (Okun 1959; Theil 1961, 1966; Suits 1962; Stekler 1968). A comprehensive NBER study initiated in 1963 resulted in a systematic collection and appraisal of annual and quarterly, public and private, judgmental and econometric forecasts of important economic aggregates and indexes as well as such events as business-cycle peaks and troughs (Zarnowitz 1967, 1972; Fels and Hinshaw 1968; Mincer 1969; Moore 1969; Cole 1969; Evans et al. 1972; Haitovsky, Treyz, and Su 1974).

In 1968, a regular quarterly survey of general economic forecasts was established at the initiative of Geoffrey Moore, then president of the ASA, to be conducted cooperatively by the NBER and the Business and Economic Statistics Section of the ASA.[5] This was the first major organized effort to build up reliable information about the potential and limitations of short-term aggregative economic forecasts, which would provide a broad base for research and improvements in this field. The ASA "agreed to carry out the surveys for a period long enough to assure accumulation of useful experience and evidence," while the National Bureau "assumed responsibility for the tabulation of forecasts, computation of error statistics and other measures, and research in evaluating the results and their analytical implications" (Zarnowitz 1968, 1–2). The cooperation was to last twenty-two years. One measure of its success is that, in 1990, the Federal Reserve Bank of Philadelphia undertook to continue the survey essentially in the same way as it was conducted by the NBER and the ASA.

## 1.2 The NBER-ASA Survey: Characteristics, Measures, and Data

### 1.2.1 Coverage

Table 1.1 identifies each of the variables covered by title, source, symbol, Commerce series number, and the form in which we use the data. During the period 1968:IV–1981:II (col. 5), direct forecasts were made for seven nominal indicators and three real indicators; also, predictions for GNP in constant dollars were derived from those for GNP in current dollars and the implicit price deflator. During the period 1981:III–1990:I (col. 6), direct forecasts were made for six nominal and eleven real variables. Seven major expenditure com-

4. Later studies of the Livingston forecasts generally used them as amended by Carlson, but many earlier studies suffer from measurement errors in the published group averages.

5. The Business and Economic Statistics Section had long been engaged in producing annual surveys of forecasts by its members.

**Table 1.1** List of Variables Covered in the NBER-ASA Quarterly Economic Outlook Surveys, 1968:IV–1981:II and 1981:III–1990:I

| Row | Variable (Symbol) (1) | Unit (R or N)[a] (2) | Source[b] (3) | Series No.[c] (4) | Period Covered 68:IV–81:II (5) | Period Covered 81:III–90:I (6) | Form[d] (7) |
|---|---|---|---|---|---|---|---|
| 1 | Gross national product (GNP) | \$bil. (N) | 1 | 200 | √ | √ | %Δ |
| 2 | GNP implicit price deflator (IPD) | b.y. = 100 (N) | 1 | 310 | √ | √ | %Δ |
| 3 | GNP in constant dollars (RGNP) | const. \$bil (R) | 1 | 50 | √ | √ | %Δ |
| 4 | Industrial production (IP) | b.y. = 100 (R) | 4 | 47 | √ | √ | %Δ |
| 5 | Unemployment rate (UR) | % (R) | 3 | 43 | √ | √ | level |
| 6 | Corporate profits after taxes (CP) | \$bil. (N) | 1 | 16 | √ | √ | %Δ |
| 7 | Plant and equipment expenditures (PE) | \$bil. (N) | 2 | 61 | √ | | %Δ |
| 8 | Private nonfarm housing starts (HS) | a.r., mil. (R) | 2 | 28 | √ | √ | level |
| 9 | Change in business inventories (CBI) | \$bil. (N) | 1 | 245 | √ | | level |
| 10 | Consumer expenditures for durable goods (CD) | \$bil. (N) | 1 | 232 | √ | | %Δ |
| 11 | National defense purchases (DEF) | \$bil. (N) | 1 | 564 | √ | | %Δ |
| 12 | Personal consumption expenditures (PCE) | const. \$bil. (R) | 1 | 231 | | √ | %Δ |

| | | | | | | |
|---|---|---|---|---|---|---|
| 13 | Nonresidential fixed invenstment (NFI) | const. $bil. (R) | 1 | 86 | √ | %Δ |
| 14 | Residential fixed investment (RFI) | const. $bil. (R) | 1 | 89 | √ | %Δ |
| 15 | Federal government purchases (FGP) | const. $bil. (R) | 1 | 263 | √ | %Δ |
| 16 | State and local government purchases (SLGP) | const. $bil. (R) | 1 | 267 | √ | %Δ |
| 17 | Change in business inventories (RCBI) | const. $bil. (R) | 1 | 30 | √ | level |
| 18 | Net exports of goods and services (NX) | const. $bil. (R) | 1 | 255 | √ | level |
| 19 | Consumer price index (CPI) | % change (N) | 3 | 320 | √ | level |
| 20 | Treasury-bill rate, 3 month (TRB) | % (N) | 4 | 114 | √ | level |
| 21 | New high-grade corporate bond yield (CBY) | % (N) | 5 | 116 | √ | level |

[a]R = real; N = nominal; b.y. = base year; a.r. = annual rate; const. $ = in constant dollars.

[b]Sources are as follows: 1 = U.S. Department of Commerce, Bureau of Economic Analysis; 2 = U.S. Department of Commerce, Bureau of the Census; 3 = U.S. Department of Labor, Bureau of Labor Statistics; 4 = Board of Governors of the Federal Reserve System; 5 = Citibank and U.S. Department of the Treasury.

[c]As listed in the *Business Conditions Digest* and the *Survey of Current Business*.

[d]As used in the computation of forecast errors. %Δ = percentage change.

ponents of real GNP, the consumer price index, the Treasury-bill rate, and the corporate bond yield were added to the list; four nominal series (expenditures for consumer durables, plant and equipment, and national defense and change in business inventories) were dropped.

The change in 1981 resulted from new initiatives undertaken by the NBER in the preceding year. A special questionnaire mailed to a long list of professional forecasters (both past and present survey participants and others) collected much useful information about the reactions to the design and uses of the NBER-ASA survey, the improvements suggested, and the assumptions and procedures favored. There was strong sentiment for expanding the survey by including several additional variables. The problem was how to comply with these wishes without either losing the essential continuity or overloading the survey and risking discouraging future participation. An advisory committee helped make the desirable changes.[6]

A large number of individuals participated in the earliest surveys, but many were not prepared to fill out a detailed questionnaire each quarter and soon dropped out. Of the more than 150 people who responded to the survey at one time or another, many did so only sporadically, and some submitted incomplete questionnaires. To exclude such occasional forecasters, we decided to use only the responses of those who answered at least ten surveys, providing information for most variables and horizons. Note that the surveys need not be consecutive; had we required long records of uninterrupted participation, few respondents would have qualified.

Table 1.2 shows how this selection was accomplished and with what results. Using the forecasts of spending on consumer durables for 1968–81, the number of respondents fell from a total of 156 to eighty-six in the sample, but the average number of surveys covered per respondent was greatly increased (e.g., doubling from eleven to twenty-two, according to the medians). The average number of respondents per survey was reduced only slightly, remaining above forty. The variability of coverage over time was lowered considerably throughout (cf. cols. 1 and 2).

The participation rates in the surveys were much smaller in 1981–90 than in 1968–81. In terms of the forecasts of real nonresidential investment, the number of respondents fell from a total of seventy-four to twenty-nine in the sample. Again, however, the selection process achieved relatively good results. The retained forecasters averaged about twenty surveys, more than double the number for all survey participants. The median number of surveys covered per respondent declined only slightly, from twenty-one to eighteen.

6. The committee was established with the support of the Business and Economic Statistics Section of the ASA and its 1980 and 1981 chairs, Arnold Zellner and George Tiao. The members included Rosanne Cole, Ray C. Fair, Edgar R. Fiedler, Albert A. Hirsch, F. Thomas Juster, Geoffrey H. Moore, George L. Perry, W. Allen Spivey, and Victor Zarnowitz. For more detail on these initiatives, see Zarnowitz (1982, 11–13).

**Table 1.2 NBER-ASA Quarterly Economic Outlook Surveys, All Forecasts and Sampled Forecasts: Selected Distributional Statistics, 1968–90 and Two Subperiods**

| | | 1968:IV–1981:II | | 1981:III–1990:I | | 1968:IV–1990:I | |
|---|---|---|---|---|---|---|---|
| Row | Statistic | All (1) | Sample (2) | All (3) | Sample (4) | All (5) | Sample (6) |
| | | Number of Surveys | | | | | |
| 1 | Total number | 51 | 51 | 35 | 35 | 86 | 86 |
| | Surveys per respondent: | | | | | | |
| 2 | Mean | 14.8 | 24.2 | 10.3 | 20.8 | 21.0 | 28.5 |
| 3 | Standard deviation | 13.0 | 10.4 | 9.9 | 7.5 | 16.1 | 13.5 |
| 4 | Median | 11 | 22 | 6 | 20 | 12 | 25 |
| 5 | Interquartile range | 21 | 18 | 14.8 | 9.5 | 26 | 21 |
| 6 | Maximum | 46 | 46 | 35 | 35 | 70 | 70 |
| 7 | Minimum | 1 | 10 | 1 | 10 | 1 | 10 |
| | | Number of Respondents | | | | | |
| 8 | Total number | 156 | 86 | 74 | 29 | 159 | 111 |
| | Respondents per survey: | | | | | | |
| 9 | Mean | 45.8 | 40.8 | 21.7 | 17.2 | 39.0 | 36.8 |
| 10 | Standard deviation | 14.5 | 11.3 | 5.9 | 3.3 | 15.9 | 14.1 |
| 11 | Median | 44 | 42 | 21 | 18 | 37 | 34.5 |
| 12 | Interquartile range | 24 | 16 | 10 | 6 | 26.2 | 22.5 |
| 13 | Maximum | 86 | 61 | 33 | 22 | 78 | 67 |
| 14 | Minimum | 22 | 20 | 10 | 9 | 12 | 12 |

*Note:* The counts refer to the forecasts one and two quarters ahead for the following variables: 1968:IV–1981:II (51 surveys), consumer expenditures for durable goods (CD); 1981:III–1990:I (35 surveys), nonresidential fixed investment (NFI); 1968:IV–1990:I (86 surveys), unemployment rate (UR). The sample includes the forecasters who participated in at least 10 surveys in terms of these observations (see row 7).

Here, too, the relevant dispersion measures were all substantially reduced (cf. cols. 3 and 4).

Finally, the sample for the total period 1968–90, based on forecasts of the unemployment rate, consists of 111 out of a total of 159 people. The coverage of surveys per respondent ranges from ten to seventy, with a mean of about twenty-eight; the corresponding figures for respondents per survey are twelve to sixty-seven and thirty-seven. Here the dispersion statistics show relatively small declines in the transition from "all" to "sample" (cf. cols. 5 and 6). All in all, the turnover among the survey participants was considerable, which should be remembered when looking at the results of our study.[7]

7. Missing observations (gaps in response) limit our ability to use these data to study such problems as the dependencies over time in the forecast errors (but see Zarnowitz 1985, sec. 3).

**Table 1.3** **Percentage Distributions of Respondents by Primary Affiliation: Four NBER-ASA Economic Outlook Surveys, 1968–80**

| | | Quarterly Surveys | | | |
|---|---|---|---|---|---|
| Row | Primary Affiliation[a] | December 1968 (1) | December 1970 (2) | November 1975 (3) | November 1980 (4) |
| 1 | Manufacturing | 39.3 | 45.6 | 21.3 | 40.0 |
| 2 | Financial institutions | 21.4 | 21.7 | 23.4 | 20.0 |
| 3 | Commercial banking | 11.9 | 6.5 | 12.8 | 13.3 |
| 4 | Other | 9.5 | 15.2 | 10.6 | 6.7 |
| 5 | Consulting and research | 11.9 | 10.9 | 23.4 | 20.0 |
| 6 | Academic | 7.1 | 4.4 | 10.6 | 6.7 |
| 7 | Government | 8.3 | 8.7 | 8.5 | 6.7 |
| 8 | Other[b] | 11.9 | 8.7 | 12.8 | 6.7 |
| 9 | Total present[c] | 100.0 (84) | 100.0 (46) | 100.0 (47) | 100.0 (30) |

[a]As reported by the participants in the given survey (those who did not respond to the question on primary affiliation are excluded).

[b]Includes a very few responses from labor union and trade association economists, but mainly "not elsewhere classified," i.e., not included in the categories listed above.

[c]Total number of respondents is listed in parentheses. The component percentages may not add up exactly to 100.0 because of rounding.

### 1.2.2 Forecasters' Affiliations and Methods

In 1968–80, the questionnaire asked the participating forecasters about their primary affiliation, but later the question was dropped. As illustrated in table 1.3, academic economists represented on average about 7 percent and government economists about 8 percent of the membership (rows 5 and 6). All other respondents, except for a few from labor unions and trade associations, came from the business world. Most of the time, manufacturing accounted for at least one-third and up to 40 percent of the participants, commercial banking and other financial institutions for one-fifth or more and consulting and research firms also for 20 percent or more in 1975–80, less in earlier years (rows 1–4).

These distributions resemble those for the universe of business forecasters as represented by the respondents to the annual economic outlook surveys of the National Association of Business Economists (NABE) in 1975–89. Here from one-third to more than 40 percent of respondents were in the industrial economy (manufacturing, energy, utilities), 25–30 percent in finance, 12 percent or more in consulting and research, 4 percent in other private services, and 6–12 percent in government and academe. The assessments of some of the NABE surveys looked for but found no systematic differences in forecasting performance between these industry groups.[8]

Another question asked regularly through 1981 concerned the relative importance assigned by survey participants to each of several items on a short list of forecasting methods or tools. Business economists use a variety of procedures to predict the major expenditure components of GNP, combine these predictions in nominal and real terms, and check and adjust the resulting forecasts for consistency with logic, theory, and the currently available information. This "informal GNP model" is an eclectic and flexible approach in which a major rule is played by the forecaster's judgment (Butler and Kavesh 1974). Over 70 percent of the NBER-ASA survey respondents reported using it, and over 50 percent on average ranked it first (table 1.4, col. 1). About one-fifth of the group favored econometric models, whether their own or outside, and one-fourth had their own econometric models (not necessarily comprehensive and first ranked). Users of outside models accounted for more than 40 percent of the early members and more than half of those in the late 1970s and early 1980s (cols. 2 and 3).

Leading indicators were employed by about 70 percent of the survey membership in 1968–70, but later that share declined to closer to 50 percent. They were ranked second by most respondents. Similar majorities referred to anticipations surveys, which generally were given lower ranks. Other methods, such as time-series models, were specified by fewer than 20 percent of the participants and preferred by about half of them (cols. 4–6).

These findings leave no doubt about one point, namely, that the listed methods were predominantly used in various combinations. Very few individuals preferred any one method to the exclusion of others. Presumably, there is a good reason for this in that 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.

There seems to be little or no systematic relation between the forecasters' rankings of the methods and the accuracy of their predictions, allowing for the difference between the targeted variables, spans, etc. This is suggested by cross-sectional (survey-by-survey) regressions of individual forecast errors on dummy variables representing the first-ranked methods as well as by comparisons of properly standardized average errors over time (Zarnowitz 1971; Su and Su 1975). The lower panel in table 1.4 (rows 5–8) presents average root mean square errors (RMSEs) for groups classified by their self-declared methodological preferences. These measures are based on a large number of individual forecasts of rates of change in nominal GNP and real GNP (RGNP),

8. We are indebted to David L. Williams, secretary-treasurer of the NABE, for help in collecting these data.

Table 1.4 **Average Ranks and Accuracy of Forecasting Methods Used in the NBER-ASA Surveys, 1968–81**

| Row | Statistic | Informal GNP Model (1) | Econometric Models: Own (2) | Econometric Models: Outside (3) | Leading Indicators (4) | Anticipations Surveys (5) | Other Methods[a] (6) |
|---|---|---|---|---|---|---|---|
| 1 | % using[b] | 75 | 24 | 48 | 62 | 57 | 16 |
| 2 | % ranking first[c] | 55 | 11 | 9 | 11 | 2 | 8 |
| 3 | % ranking second | 13 | 7 | 15 | 29 | 21 | 4 |
| 4 | % ranking lower[d] | 6 | 7 | 25 | 22 | 35 | 4 |
| | | Average Root Mean Square Error[e] | | | | | |
| 5 | GNP,[f] % change | .96 | 1.09 | .89 | 1.00 | .99 | 1.15 |
| 6 | RGNP, % change | 1.14 | 1.25 | 1.05 | 1.24 | 1.22 | 1.27 |
| 7 | IPD, % change | .71 | .76 | .72 | .79 | .85 | .83 |
| 8 | UR, level | .58 | .66 | .52 | .62 | .71 | .59 |

[a]A "write-in" response but often not specified.

[b]Based on seven surveys 1968:IV–1970:II (496 replies), six surveys 1974:I–1975:II (308 replies), and six surveys 1980:I–1981:II (187 replies). 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 covered varied from 24 to 83. The averages are weighted according to the numbers of the replies.

[c]Most important.

[d]Ranks third to sixth (least important).

[e]According to first-ranked method (ties for the first rank are not included). Refers to 79 individuals who participated in at least 12 of the 46 quarterly surveys in the period from 1968:IV through 1980:I. For more detail, see Zarnowitz (1983).

[f]Symbols are defined in table 1.1.

IPD (implicit price deflator) inflation, and the levels of the unemployment rate; they omit occasional forecasters and aggregate across predictions for the current quarter and three quarters ahead. The differences between the RMSEs are generally small and of uncertain significance.[9]

### 1.2.3 Basic Measures of Error in Forecasts of Changes and Levels

For series with upward trends, for example, GNP in current and constant dollars and the implicit price deflator, the most relevant forecasts are those of percentage change. Let the current survey quarter and the four quarters that follow be denoted by $t = 1, \ldots, 5$, respectively. The most recent quarter for which data are available precedes the date of the survey ($t = 0$). Then the predicted *average* changes refer to the spans 0–1, 0–2, . . . , 0–5, and the implied *marginal* (or intraforecast) changes refer to the spans 0–1, 1–2, . . . , 4–5.

For approximately stationary series such as the unemployment rate, real inventory investment, and real net exports, the most relevant forecasts are those of levels in the original units. They refer to quarters 1, . . . , 5.

Our data consist of more than 17,000 individual time series of forecasts defined by source, variable, and horizon. For example, for 1968–90, there are 111 respondents in our sample, reporting on seven variables over five spans each, yielding 3,885 series ($= 111 \times 7 \times 5$; however, consideration of four marginal changes for five of these variables adds another subset of 2,220 series). The tables presented below record the distributions of the summary measures of error across these individual series for each variable, period, and horizon covered. We distinguish three measures—the mean error (ME), the mean absolute error (MAE), and the root mean square error (RMSE)—and compute several location and dispersion statistics for each. These statistics include means, standard deviations, medians, interquartile ranges, skewness, and kurtosis (denoted by M, SD, MD, IQR, SK, and KU, respectively). Not all the detail of this compilation can be presented here, of course, but it is available for purposes of verification and further research.

### 1.2.4 Data Revisions and Forecast Accuracy

Some of the variables covered by the surveys, such as the consumer price index and the interest rates, are subject to few or no revisions. Others, notably the aggregates and indexes taken from the national income and product accounts (NIPAs), are revised frequently, and some of the revisions are large. An old but still controversial issue is which revision or vintage of such data should be used in evaluating the accuracy of forecasts. The preliminary figures are most closely related to the latest figures that were available to the

9. Most of these differences actually disappear when rounded off to one decimal point. Providing detail by span of forecast and for some other variables would not alter the picture significantly (see Zarnowitz 1983, 84–85). However, it is probably worth noting that the group ranking first the outside econometric models had the smallest average RMSEs for most variables (col. 3). This group included large companies using well-known econometric service bureaus as well as their own staffs of professional economists.

forecasters, but they may themselves be partly predictions or "guesstimates" and may seriously deviate from "the truth," as represented by the last revision of the data. On the other hand, the final data may be issued years after the forecast was made and may incorporate major benchmark revisions. That the forecasters should be responsible for predicting all measurement errors to be corrected by such revisions is surely questionable.

Appraisals of forecasts differ: Some are based on early data (e.g., Zarnowitz 1967), others on late data, generally prebenchmark revisions (e.g., McNees 1979; Zarnowitz 1985). Judgmental forecasts that rely heavily on recent preliminary figures may look best when compared with early data; econometric model forecasts that incorporate long series of revised data may be more favored by evaluations using later vintages.

For the NBER-ASA percentage change forecasts of GNP, RGNP, and IPD, table 1.5 shows the MAEs and RMSEs obtained by comparisons with fifteen-day, forty-five-day, early July, and late July data. In general, and with exceptions, the errors tend to increase monotonically the more revised the data are. However, the differences between the successive error measures in each segment and column of the table are relatively small, typically less than 0.1 percent. This is fortunate because it suggests that the choice of which vintage of the data to use may not be so critical. Even so, larger differences may occur in particular subperiods and offset each other over the total period covered. Our results certainly do not detract from the importance of measurement errors in the forecasting context, which has been demonstrated to be large (Cole 1969).

To save space and avoid relying on the extremes of either very preliminary or repeatedly revised data, we shall henceforth use the forty-five-day estimate in most of our text references and all tabular presentations. But no single data vintage is an optimal standard here; the choice of any is inevitably more or less arbitrary and too restrictive.

## 1.3 Forecasts of Nominal and Real GNP Growth and Inflation

### 1.3.1 Graphic Comparisons of Predictions

A convenient way to relate the distributions of survey forecasts and the actual data visually is to plot the former in the form of box diagrams and the latter as a continuous series, quarter by quarter, to common scales. Figures 1.1–1.3 apply this device to predictions of nominal and real GNP growth and IPD inflation rates. There is one graph for each variable and horizon. The midpoint of each box marks the location of the group's mean forecast; the top and bottom mark the mean plus or minus one standard deviation. A longer vertical line bisects each box and connects the highest and lowest forecasts recorded on the same occasion. A heavy curve superimposed on the array of the boxes and vertical lines represents the actual outcomes (forty-five-day estimates).

The graphs make it clear that the curves cross most of the boxes. This

**Table 1.5 Mean Absolute Errors and Root Mean Square Errors of Forecasts of Nominal and Real GNP Growth and Inflation: Comparisons with Different Vintages of Target Data, 1968–90**

| | | Mean Absolute Errors by Span (Qs)[b] | | | | | Root Mean Square Errors by Span (Qs)[c] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Row | Vintage of Actual Data[a] | 0–1 (1) | 0–2 (2) | 0–3 (3) | 0–4 (4) | 0–5 (5) | 0–1 (6) | 0–2 (7) | 0–3 (8) | 0–4 (9) | 0–5 (10) |
| | | Gross National Product (GNP) | | | | | | | | | |
| 1 | 15-day | .59 | 1.08 | 1.55 | 1.92 | 2.36 | .77 | 1.41 | 2.03 | 2.54 | 3.13 |
| 2 | 45-day | .62 | 1.12 | 1.60 | 1.99 | 2.48 | .86 | 1.45 | 2.07 | 2.58 | 3.20 |
| 3 | Early July | .65 | 1.15 | 1.65 | 2.02 | 2.54 | .85 | 1.48 | 2.10 | 2.60 | 3.26 |
| 4 | Late July | .69 | 1.17 | 1.66 | 2.03 | 2.52 | .89 | 1.50 | 2.10 | 2.60 | 3.23 |
| | | Gross National Product in Constant Dollars (RGNP) | | | | | | | | | |
| 5 | 15-day | .61 | 1.06 | 1.51 | 1.96 | 2.44 | .81 | 1.40 | 2.04 | 2.70 | 3.35 |
| 6 | 45-day | .64 | 1.09 | 1.56 | 2.00 | 2.47 | .85 | 1.44 | 2.08 | 2.74 | 3.38 |
| 7 | Early July | .67 | 1.09 | 1.57 | 1.99 | 2.46 | .88 | 1.44 | 2.07 | 2.69 | 3.33 |
| 8 | Late July | .68 | 1.11 | 1.58 | 2.01 | 2.48 | .90 | 1.44 | 2.05 | 2.66 | 3.30 |
| | | Implicit Price Deflator (IPD) | | | | | | | | | |
| 9 | 15-day | .40 | .71 | 1.07 | 1.49 | 1.98 | .50 | .92 | 1.37 | 1.92 | 2.56 |
| 10 | 45-day | .42 | .77 | 1.16 | 1.63 | 2.14 | .54 | .99 | 1.50 | 2.10 | 2.79 |
| 11 | Early July | .42 | .77 | 1.18 | 1.66 | 2.17 | .53 | .99 | 1.52 | 2.14 | 2.83 |
| 12 | Late July | .41 | .79 | 1.21 | 1.70 | 2.21 | .53 | .99 | 1.53 | 2.16 | 2.84 |

[a] 15-day: preliminary data released in the month following the target quarter of the forecast. 45-day: revised data released a month later. Early July: generally first July revision; where this is not available, the preceding revision. Late July: generally second July revision; where this is not available, the preceding revision.

[b] Mean of the MAEs of the individual forecasts, where MAE $= 1/N \Sigma \mid E_t \mid$; $E_t = P_t - A_t$; $P_t$ = predicted value; $A_t$ = actual value of the given vintage. The average errors refer to percentage changes from quarter $t - 1$ (0) to quarters $t$, $t + 1$, $t + 2$, $t + 3$, and $t + 4$ (1, 2, 3, and 4), respectively, where $t$ refers to the quarterly date of the survey. Thus, 0–1 denotes the change from quarter $t - 1$ to quarter $t$, 0–2 denotes the change from quarter $t - 1$ to quarter $t + 1$, etc. All measures refer to percentage change errors and are given in percentages.

[c] Mean of the RMSEs of the individual forecasts, where RMSE $= \sqrt{1/n \, \Sigma \, (P_t - A_t)}$.

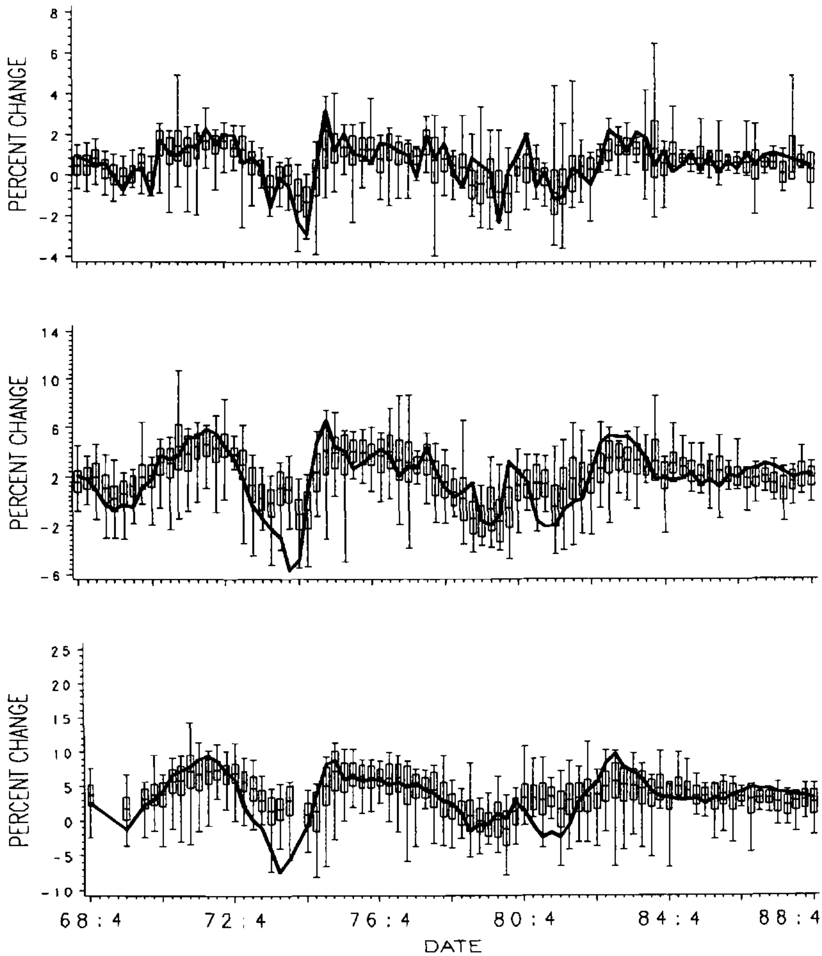


**Fig. 1.1 Forecast distributions and actual values of percentage changes in real GNP, three horizons, 1968:IV–1990:I**

H = high; M = mean; SD = standard deviation; L = low.

means that the realizations fall within one standard deviation of the mean or "consensus" predictions most of the time. However, some large declines in actual values are widely missed or underestimated, such mis- or underestimations showing up as boxes lying conspicuously above the troughs or valleys in the curves. Similarly, widespread underpredictions of some large actual rises

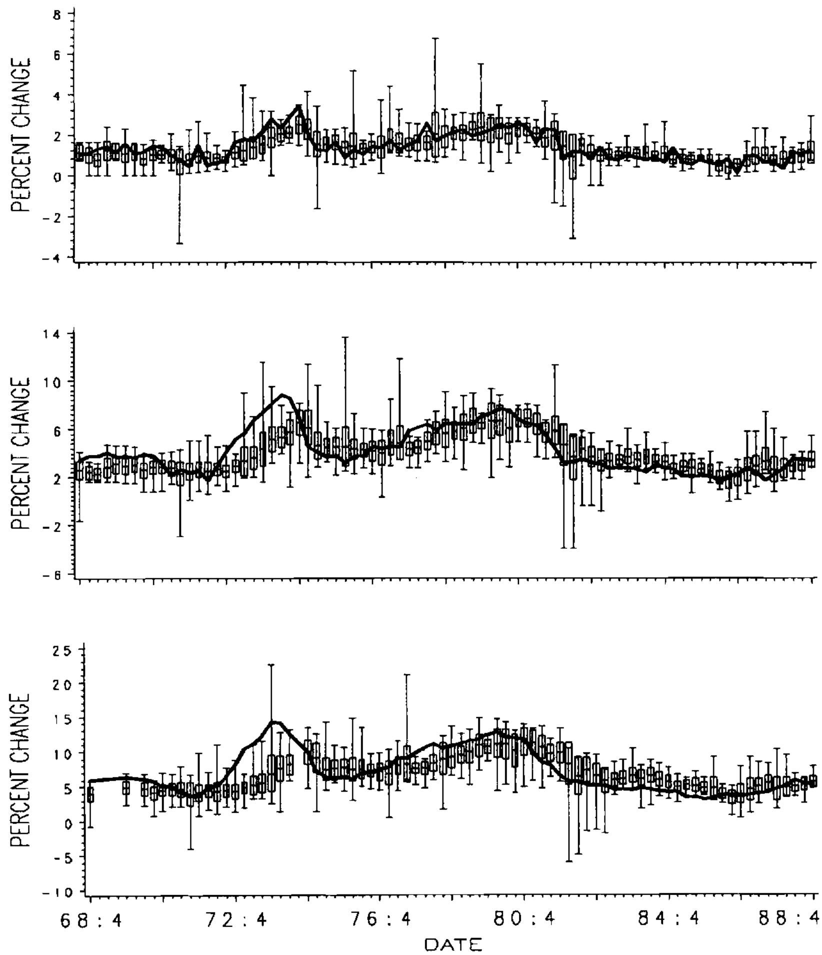


**Fig. 1.2 Forecast distributions and actual values of percentage changes in IPD, three horizons, 1968:IV–1990:I**

show up as boxes situated below the local peaks for concentrations of high values. Occasionally, the actual outcome would be missed by all respondents to the survey, as seen in instances where the entire vertical line of forecasts lies above or below the curve.

These errors are clearly associated with business cycles. Figure 1.1 shows clusters of large overestimates of real GNP growth in all major slowdowns and recessions covered: 1969–70, 1973–74, 1981–82, and 1985–86. It also shows clusters of large underestimation errors in all recoveries and booms: 1972, 1975, late 1980, 1983–84, and 1987. So overprediction of growth occurs mainly when the economy weakens and declines, underprediction when it rises strongly. Both types of error can be seen as particularly pronounced

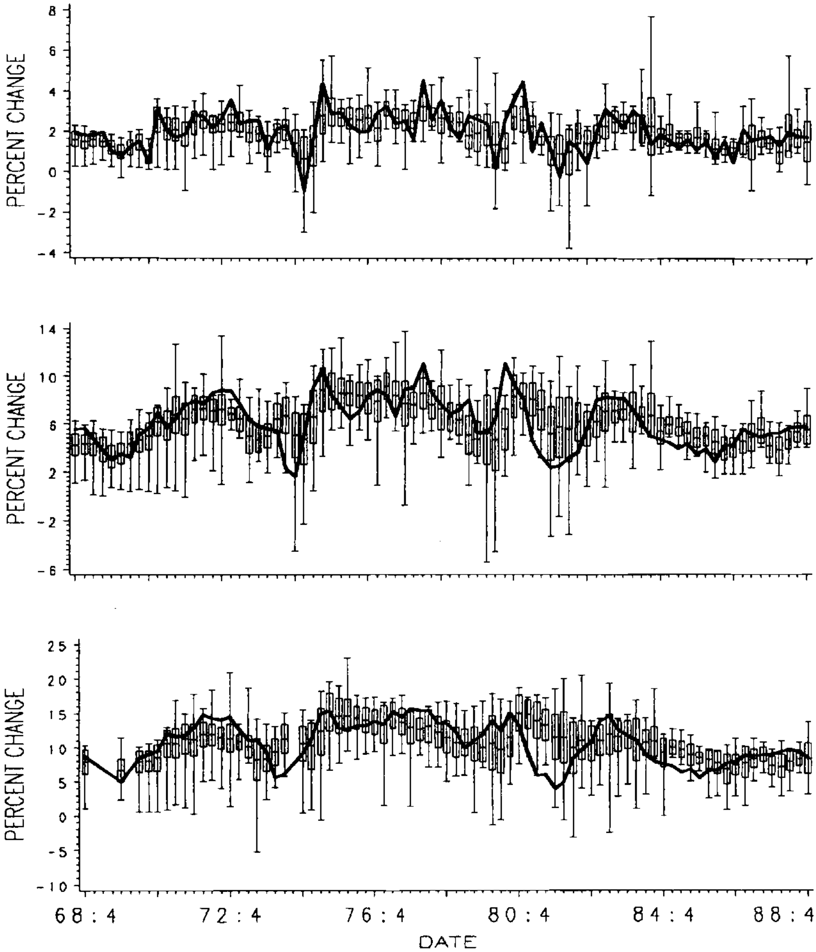


**Fig. 1.3 Forecast distributions and actual values of percentage changes in nominal GNP, three horizons, 1968:IV–1990:I**

and persistent in forecasts with longer spans. Overall, the errors of overprediction in bad times tended to be larger than those of underprediction in good times.

Figure 1.2 shows that inflation was at times widely underpredicted in 1969–71, even though it was then fairly stable. In 1973–74, a period of supply shocks and deepening recession, inflation rose sharply and was greatly underestimated by most survey participants. Here the curves can be seen to rise above most of the boxes and even to peak above the highest forecasts for the longer horizons. The same tendency to underpredict also prevailed in 1976–80, although in somewhat weaker form. In this period, inflation rose more gradually, while the economy first expanded vigorously and then, in 1979–80, experienced another oil shock, a slowdown, and a short recession. In between, during the recovery of 1975–76, inflation decreased markedly and was mostly overestimated. Another, much longer disinflation occurred in 1981–85, a phase that followed the shifts to a tight monetary policy in late 1979 and included the severe 1981–82 recession and then a strong recovery. Here again most forecasters are observed to overpredict inflation. Finally, in 1986–89, inflation, which began to drift upward, was generally well predicted most of the time (except in the mid-quarter of 1987, when it dipped suddenly and was overestimated).

In sum, there is also a cyclical pattern to the errors of inflation forecasts. Accelerated inflation was associated predominantly with underprediction, disinflation with overprediction errors.

Figure 1.3, which compares the forecast distributions and actual values for nominal GNP growth rates, shows a broad family resemblance to the corresponding graphs for real GNP growth in figure 1.1. For example, both nominal and real growth tended to be underpredicted in such boom years as 1972 and 1983 and overpredicted in such recession years as 1974 and 1981–82. But inflation expectations and their relation to real growth forecasts are also important here. Predictions of nominal GNP are often helped by inverse correlations between the changes in IPD and RGNP and the associated offsets between the forecast errors for the two variables.[10] Thus, in the inflationary recession of 1973–74 associated with the first occurrence of major supply and oil shocks, real growth was overpredicted and inflation underpredicted. The reverse combination of too low RGNP and too high IPD forecasts can be observed in the recoveries of 1974 and 1983–84. However, there are also episodes of positive correlation; for example, in 1981–82, both real growth and inflation were overpredicted, which resulted in nominal growth forecasts that turned out much too high.

### 1.3.2 Distributions of Summary Measures of Error

Table 1.6 presents the statistics on the distributions of the mean errors in the sampled NBER-ASA survey forecasts of GNP, RGNP, and IPD. For the forecasts of average changes in GNP, the means are all negative, but the corresponding medians have mixed signs. The averages for the marginal change errors are predominantly positive. The dispersion measures (SD and IQR) are very large relative to the averages. Thus, these statistics (rows 1–4) fail to show clearly any dominant under- or overprediction bias. Similar observations can be made about the real GNP forecasts (rows 7–10). However, underestimation errors definitely prevail in the inflation (IPD) forecasts. Here all the averages, M and MD, are negative, and the relative size of the corresponding AD and IQR figures is less.

10. This has been noted before, in Zarnowitz (1979, 15).

**Table 1.6 Distribution of Mean Errors in Individual Forecasts of Nominal and Real GNP Growth and Inflation, 1968–90**

| | | Average Errors by Span (Qs) | | | | | Marginal Errors by Span (Qs) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 0–1 | 0–2 | 0–3 | 0–4 | 0–5 | 1–2 | 2–3 | 3–4 | 4–5 |
| Row | Statistic | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| | | Gross National Product (GNP) | | | | | | | | |
| 1 | Mean (M) | −.09 | −.11 | −.11 | −.14 | −.30 | −.01 | .01 | .04 | .08 |
| 2 | Standard deviation (SD) | .26 | .56 | .91 | 1.19 | 1.61 | .31 | .38 | .36 | .37 |
| 3 | Median (MD) | −.05 | −.06 | .04 | .15 | .00 | .01 | .06 | .09 | .13 |
| 4 | Interquartile range (IQR) | .24 | .48 | .72 | .91 | 1.30 | .27 | .29 | .31 | .31 |
| 5 | Skewness (SK) | −1.94 | −1.57 | −1.60 | −2.65 | −2.78 | −.55 | −1.41 | −1.64 | −1.92 |
| 6 | Kurtosis (KU) | 12.14 | 12.42 | 11.90 | 12.81 | 13.80 | 10.89 | 8.49 | 8.50 | 8.66 |
| | | Gross National Product in Constant Dollars (RGNP) | | | | | | | | |
| 7 | Mean (M) | −.01 | .09 | .25 | .45 | .48 | .10 | .16 | .22 | .28 |
| 8 | Standard deviation (SD) | .24 | .48 | .77 | 1.06 | 1.39 | .29 | .35 | .34 | .43 |
| 9 | Median (MD) | −.00 | .13 | .34 | .62 | .64 | .10 | .20 | .27 | .29 |
| 10 | Interquartile range (IQR) | .27 | .46 | .69 | .98 | 1.20 | .28 | .36 | .38 | .51 |
| 11 | Skewness (SK) | −1.30 | −1.58 | −1.84 | −2.04 | −2.06 | −1.17 | −1.75 | −1.34 | −.89 |
| 12 | Kurtosis (KU) | 4.64 | 6.76 | 7.29 | 7.57 | 8.13 | 4.78 | 6.63 | 4.24 | 1.97 |
| | | Implicit Price Deflator (IPD) | | | | | | | | |
| 13 | Mean (M) | −.07 | −.19 | −.36 | −.57 | −.65 | −.12 | −.15 | −.19 | −.21 |
| 14 | Standard deviation (SD) | .16 | .34 | .57 | .83 | 1.24 | .20 | .25 | .27 | .35 |
| 15 | Median (MD) | −.07 | −.17 | −.34 | −.54 | −.74 | −.11 | −.14 | −.16 | −.21 |
| 16 | Interquartile range (IQR) | .15 | .39 | .75 | 1.09 | 1.79 | .27 | .35 | .36 | .52 |
| 17 | Skewness (SK) | .06 | .32 | .09 | .14 | −.06 | .28 | −.10 | .26 | −.04 |
| 18 | Kurtosis (KU) | 1.35 | .76 | .50 | .92 | −.11 | .42 | .36 | 1.91 | −.24 |

*Note:* Columns 1–5 refer to the errors in forecasts of average changes; cols. 6–9 refer to the errors in forecast of marginal changes (for 0–1, the average and marginal changes are the same). ME, SD, MD, and IQR (rows 1–4, 7–10, and 13–16) are in percentage points; entries for SK and KU (rows 5–6, 11–12, and 17–18) are dimensionless ratios. IQR = $Q_3 - Q_1$ is the difference, third quartile minus first quartile of the distribution (where MD = $Q_2$). SK = $\mu_3/\sigma^3$ is the ratio of the third moment around the mean to the third power of the standard deviation SD = $\sigma$. KU = $\mu_4/\sigma^4$ is the ratio of the fourth moment around the mean to the fourth power of SD.

The M and MD statistics tend to increase monotonically in absolute value with the length of the span, strongly for the forecasts of average change, less so for those of the marginal change. For each of three variables, the SD and IQR statistics tend to be much larger the longer the span and the more remote the forecast target (cf. rows 1–4, 7–10, and 13–16).

There is evidence that the distributions for GNP and RGNP are skewed to the left (i.e., SK $< 0$), with medians larger than the means. For IPD, SK is very small throughout, and M and MD are very close (cf. rows 5, 11, and 17).

The distributions for GNP and RGNP show larger values for kurtosis, indicating the presence of long, thick tails (for the normal distribution, KU $= 3$). Again, the situation is very different for IPD, where the KU statistics are very low (cf. rows 6, 12, and 18).

Tables 1.7 and 1.8, each of which has the same format as table 1.6, show the distribution statistics for the mean absolute errors and the root mean square errors, respectively. The RMSEs are, of course, larger than the corresponding MAEs, and the statistics in table 1.8 are generally larger than their counterparts in table 1.7 (e.g., they are on average about 30–60 percent higher for the GNP measures). Otherwise, the two sets have very similar characteristics, which can be summed up as follows.

For both the MAEs and the RMSEs of the individual forecasts, the means and medians increase with the span regularly, strongly for the average changes, less so for the marginal changes. The main reason is that errors cumulate over time, but it is also true that the more distant target quarters are predicted somewhat less accurately than the near ones. The dispersion statistics SD and IQR also increase as the forecast horizon lengthens, except for the marginal IPD errors.

SK is greater than zero everywhere here, and the SK statistics are generally large for GNP and RGNP but small for IPD. Consistently, the MDs tend to be smaller than the MEs. The distributions tend to be skewed to the right.

Several of the KU statistics for GNP and RGNP are quite large. Little kurtosis is observed in the IPD forecasts, except for the shortest ones.

We conclude that the survey respondents tended to underestimate inflation but not (or, in any event, much less) the nominal and real GNP growth rates. The IPD forecast distributions were more nearly symmetrical and had fewer outliers than the distributions for GNP and RGNP.

### 1.3.3 Individual versus Group Mean Forecasts

Combining corresponding forecasts that come from different sources or use different techniques tends to produce significant gains in accuracy. This is by now well known from many studies, including some based on the NBER-ASA surveys.[11] In what follows, we extend and update the evidence on this point.

11. See Zarnowitz (1984), which uses the data for 1968–79. An early demonstration that simple averaging can reduce forecast errors is in Zarnowitz (1967, 123–26). For a survey of the literature, see Clemen (1989).

**Table 1.7 Distribution of Mean Absolute Errors in Individual Forecasts of Nominal and Real GNP Growth and Inflation, 1968–90**

| | | Average Errors by Span (Qs) | | | | | Marginal Errors by Span (Qs) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Row | Statistic | 0–1 (1) | 0–2 (2) | 0–3 (3) | 0–4 (4) | 0–5 (5) | 1–2 (6) | 2–3 (7) | 3–4 (8) | 4–5 (9) |
| | | Gross National Product (GNP) | | | | | | | | |
| 1 | Mean (M) | .62 | 1.12 | 1.60 | 1.99 | 2.48 | .76 | .84 | .85 | .88 |
| 2 | Standard deviation (SD) | .23 | .43 | .64 | .81 | 1.12 | .22 | .23 | .23 | .26 |
| 3 | Median (MD) | .56 | 1.02 | 1.49 | 1.84 | 2.31 | .73 | .81 | .82 | .86 |
| 4 | Interquartile range (IQR) | .28 | .39 | .41 | .61 | .98 | .20 | .22 | .26 | .28 |
| 5 | Skewness (SK) | 1.69 | 2.38 | 3.11 | 3.51 | 3.85 | 1.61 | 1.69 | 1.16 | 1.11 |
| 6 | Kurtosis (KU) | 4.18 | 8.21 | 13.31 | 18.79 | 22.53 | 4.15 | 5.44 | 2.46 | 2.21 |
| | | Gross National Product in Constant Dollars (RGNP) | | | | | | | | |
| 7 | Mean (M) | .64 | 1.09 | 1.56 | 2.00 | 2.47 | .78 | .85 | .93 | .96 |
| 8 | Standard deviation (SD) | .23 | .37 | .51 | .63 | .84 | .20 | .20 | .24 | .26 |
| 9 | Median (MD) | .59 | 1.00 | 1.41 | 1.82 | 2.26 | .76 | .82 | .92 | .94 |
| 10 | Interquartile range (IQR) | .22 | .33 | .46 | .63 | .89 | .29 | .29 | .34 | .33 |
| 11 | Skewness (SK) | 1.59 | 1.77 | 1.82 | 1.76 | 1.92 | 1.09 | .70 | .30 | .33 |
| 12 | Kurtosis (KU) | 2.89 | 4.12 | 4.12 | 4.13 | 6.12 | 2.20 | .87 | −.18 | .36 |
| | | Implicit Price Deflator (IPD) | | | | | | | | |
| 13 | Mean (M) | .42 | .77 | 1.16 | 1.63 | 2.14 | .50 | .56 | .61 | .65 |
| 14 | Standard deviation (SD) | .13 | .23 | .35 | .45 | .59 | .12 | .13 | .16 | .18 |
| 15 | Median (MD) | .38 | .72 | 1.08 | 1.55 | 2.07 | .49 | .54 | .57 | .62 |
| 16 | Interquartile range (IQR) | .14 | .23 | .47 | .51 | .69 | .15 | .18 | .16 | .19 |
| 17 | Skewness (SK) | 1.84 | 1.29 | 1.11 | .71 | .64 | .85 | 1.17 | 1.11 | .98 |
| 18 | Kurtosis (KU) | 4.53 | 2.51 | 2.21 | .85 | .84 | 1.36 | 1.25 | 2.20 | 1.76 |

*Note:* See table 1.6.

**Table 1.8** **Distribution of Root Mean Square Errors in Individual Forecasts of Nominal and Real GNP Growth and Inflation, 1968–90**

| | | Average Errors by Span (Qs) | | | | | Marginal Errors by Span (Qs) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Row | Statistic | 0–1 (1) | 0–2 (2) | 0–3 (3) | 0–4 (4) | 0–5 (5) | 1–2 (6) | 2–3 (7) | 3–4 (8) | 4–5 (9) |
| | | Gross National Product (GNP) | | | | | | | | |
| 1 | Mean (M) | .81 | 1.45 | 2.07 | 2.58 | 3.20 | 1.02 | 1.12 | 1.13 | 1.19 |
| 2 | Standard deviation (SD) | .30 | .55 | .76 | .92 | 1.25 | .31 | .30 | .30 | .38 |
| 3 | Median (MD) | .74 | 1.33 | 1.93 | 2.45 | 3.06 | .97 | 1.10 | 1.11 | 1.14 |
| 4 | Interquartile range (IQR) | .38 | .47 | .61 | .69 | 1.14 | .25 | .29 | .39 | .40 |
| 5 | Skewness (SK) | 1.20 | 2.17 | 2.36 | 2.34 | 2.43 | 1.63 | 1.34 | .55 | 1.51 |
| 6 | Kurtosis (KU) | 1.52 | 6.48 | 7.32 | 9.47 | 10.59 | 3.97 | 3.65 | .06 | 4.61 |
| | | Gross National Product in Constant Dollars (RGNP) | | | | | | | | |
| 7 | Mean (M) | .85 | 1.44 | 2.08 | 2.74 | 3.38 | 1.05 | 1.16 | 1.27 | 1.32 |
| 8 | Standard deviation (SD) | .35 | .49 | .67 | .82 | 1.04 | .31 | .28 | .33 | .38 |
| 9 | Median (MD) | .77 | 1.32 | 1.90 | 2.57 | 3.12 | .98 | 1.15 | 1.26 | 1.31 |
| 10 | Interquartile range (IQR) | .34 | .54 | .80 | .83 | 1.18 | .38 | .39 | .50 | .51 |
| 11 | Skewness (SK) | 1.78 | 1.54 | 1.54 | 1.30 | 1.22 | 1.37 | .47 | .18 | .46 |
| 12 | Kurtosis (KU) | 4.02 | 3.42 | 3.29 | 2.04 | 1.90 | 3.08 | .20 | −.11 | .86 |
| | | Implicit Price Deflator (IPD) | | | | | | | | |
| 13 | Mean (M) | .54 | .99 | 1.50 | 2.10 | 2.79 | .65 | .72 | .79 | .85 |
| 14 | Standard deviation (SD) | .19 | .32 | .43 | .57 | .75 | .20 | .18 | .24 | .25 |
| 15 | Median (MD) | .47 | .91 | 1.43 | 2.06 | 2.79 | .61 | .70 | .74 | .81 |
| 16 | Interquartile range (IQR) | .18 | .30 | .55 | .79 | 1.08 | .22 | .24 | .22 | .33 |
| 17 | Skewness (SK) | 1.82 | 1.44 | .89 | .54 | .18 | 1.44 | .94 | 1.51 | 1.21 |
| 18 | Kurtosis (KU) | 3.35 | 2.78 | 1.15 | 1.06 | −.04 | 3.00 | 1.58 | 4.19 | 3.64 |

*Note:* See table 1.6.

Averaging all predictions in each survey for a given variable and horizon results in a time series of group mean (or median) forecasts. These are often called "consensus" forecasts, whether or not there is much actual consensus among the respondents. The group mean predictions based on our GNP, RGNP, and IPD sample forecasts have considerably smaller errors than the average individual respondent, as shown by comparisons of the ME, MEA, and RMSE entries in table 1.9 (cf. cols. 1 and 4, 2 and 5, and 3 and 6). The absolute or squared errors tend to increase with the span of forecast for both individuals and group means, but less so for the latter.

For each individual time series of forecasts, a series of group mean forecasts has been computed with strictly matching coverage in terms of the survey dates and target characteristics. Table 1.10 shows the locational statistics for the distributions of the ratios of the individual RMSEs to the corresponding group RMSEs. These measures indicate that the group mean forecasts were more accurate than about 75 percent of the sampled respondents' forecasts. Thus, most of the first- or lower-quartile ($Q_1$) ratios are close to one (but some for RGNP are lower), most of the median ($Q_2$) ratios are 1.1–1.2, and most of the third- or upper-quartile ($Q_3$) ratios are 1.3–1.5 (cf. rows 2–4, 7–9, and 12–14). These distributions are bounded from below (any ratio greater than zero) and are heavily skewed to the right (e.g., the entries for the best forecasts in table 1.10 are 0.5–0.9, those for the worst 3–7).

Unlike their numerators and denominators, the ratios of the individual to the group RMSEs do not depend systematically on the length of the forecast or distance to the target quarter. Also, the diversity of the individual forecasts by source, variable, and horizon is greatly reduced by the normalization with the group means. Thus, the ratios for the same quartiles are not very different for GNP, RGNP, and IPD.

### 1.3.4 Some Overall Accuracy and Variability Measures

The preceding tables offer some insight into the structure of errors calculated from the survey forecasts, but not into their relative levels. The latter will be assessed by comparisons with benchmark predictions from time-series models selected to fit the characteristics of the variables concerned and with forecasts from other sources. But, first, we take a quick look at the average values of the outcomes for the target series so as to gain some idea about the orders of magnitude involved.

Columns 7–9 in table 1.9 show, successively, the means, standard deviations, and root mean square values (RMSVs) of the actual percentage changes in the targeted variables. The absolute values of the average errors in the individual forecasts and, a fortiori, in the group mean forecasts are generally very small compared with the average actual changes, particularly for GNP and IPD (cf. cols. 1 and 7). The average RMSEs of the individual forecasts are about 30–37 percent of the RMSVs for the nominal GNP growth and inflation and 68–72 percent of the RMSVs for the real GNP growth rates (cf. cols.

**Table 1.9** Individual and Group Mean Forecasts and Actual Values of Nominal and Real GNP Growth and Inflation: Selected Statistics on Accuracy and Variability, 1968–90

| | | Individual Forecasts[a] | | | Group Mean Forecasts[b] | | | Actual Values (%Δ)[c] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Row | Span (Qs) | M (1) | MAE (2) | RMSE (3) | M (4) | MAE (5) | RMSE (6) | M (7) | SD (8) | RMSV (9) |
| | | Gross National Product (GNP) | | | | | | | | |
| 1 | 0–1 | −.09 | .62 | .81 | −.08 | .49 | .64 | 1.98 | .96 | 2.20 |
| 2 | 0–2 | −.11 | 1.12 | 1.45 | −.10 | .84 | 1.11 | 4.00 | 1.00 | 4.31 |
| 3 | 0–3 | −.11 | 1.60 | 2.07 | −.07 | 1.22 | 1.61 | 6.07 | 2.14 | 6.44 |
| 4 | 0–4 | −.14 | 1.99 | 2.58 | −.02 | 1.56 | 2.06 | 8.20 | 2.63 | 8.61 |
| 5 | 0–5 | −.30 | 2.48 | 3.20 | −.09 | 1.91 | 2.51 | 10.38 | 3.12 | 10.84 |
| | | Gross National Product in Constant Dollars (RGNP) | | | | | | | | |
| 6 | 0–1 | −.01 | .64 | .85 | −.02 | .50 | .64 | .61 | 1.03 | 1.20 |
| 7 | 0–2 | .09 | 1.09 | 1.44 | .02 | .83 | 1.11 | 1.23 | 1.77 | 2.16 |
| 8 | 0–3 | .25 | 1.56 | 2.08 | .16 | 1.17 | 1.61 | 1.86 | 2.40 | 3.04 |
| 9 | 0–4 | .45 | 2.00 | 2.74 | .33 | 1.42 | 2.05 | 2.50 | 2.95 | 3.87 |
| 10 | 0–5 | .48 | 2.47 | 3.38 | .40 | 1.70 | 2.47 | 3.15 | 3.45 | 4.67 |
| | | Implicit Price Deflator (IPD) | | | | | | | | |
| 11 | 0–1 | −.07 | .42 | .54 | −.04 | .28 | .35 | 1.36 | .65 | 1.51 |
| 12 | 0–2 | −.19 | .77 | .99 | −.12 | .55 | .70 | 2.74 | 1.25 | 3.01 |
| 13 | 0–3 | −.36 | 1.16 | 1.50 | −.22 | .84 | 1.13 | 4.16 | 1.84 | 4.55 |
| 14 | 0–4 | −.57 | 1.63 | 2.10 | −.34 | 1.21 | 1.64 | 5.60 | 2.43 | 6.10 |
| 15 | 0–5 | −.65 | 2.14 | 2.79 | −.37 | 1.63 | 2.23 | 7.08 | 3.03 | 7.70 |

*Note:* On the symbols used, see previous tables and the text.

[a]Means of the corresponding statistics for individual forecasts (as shown in tables 1.6–1.8, rows 1, 7, and 13, cols. 1–5).

[b]Survey-by-survey "consensus" forecasts based on the sampled data, as explained in the text.

[c]45-day estimates, as used in tables 1.6–1.8. RMSV = root mean square value computed as $\sqrt{[(\mathrm{ME})^2 + (\mathrm{SD})^2]}$.

**Table 1.10 Individual-to-Group Mean Ratios of Root Mean Square Errors: Selected Distributional Statistics for Forecasts of Nominal and Real GNP Growth and Inflation, 1968–90**

| | | Average Errors by Span (Qs) | | | | | Marginal Errors by Span (Qs) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Row | Statistics | 0–1 (1) | 0–2 (2) | 0–3 (3) | 0–4 (4) | 0–5 (5) | 1–2 (6) | 2–3 (7) | 3–4 (8) | 4–5 (9) |
| | | Gross National Product (GNP) | | | | | | | | |
| 1 | Lowest-error forecast (MIN) | .84 | .72 | .88 | .79 | .78 | .85 | .85 | .82 | .68 |
| 2 | First quartile ($Q_1$) | 1.10 | 1.07 | 1.07 | 1.08 | 1.03 | 1.05 | 1.07 | 1.06 | 1.07 |
| 3 | Median (MD) | 1.22 | 1.23 | 1.19 | 1.20 | 1.18 | 1.15 | 1.11 | 1.13 | 1.14 |
| 4 | Third quartile ($Q_3$) | 1.62 | 1.46 | 1.38 | 1.36 | 1.43 | 1.33 | 1.25 | 1.28 | 1.32 |
| 5 | Highest-error forecast (MAX) | 7.34 | 5.35 | 5.90 | 5.78 | 5.57 | 3.12 | 2.88 | 2.56 | 4.26 |
| | | Gross National Product in Constant Dollars (RGNP) | | | | | | | | |
| 6 | Lowest-error forecast (MIN) | .82 | .63 | .77 | .82 | .82 | .76 | .87 | .86 | .83 |
| 7 | First quartile ($Q_1$) | 1.11 | 1.09 | 1.04 | 1.06 | 1.06 | 1.06 | 1.06 | 1.04 | 1.03 |
| 8 | Median (MD) | 1.30 | 1.20 | 1.17 | 1.14 | 1.14 | 1.17 | 1.14 | 1.13 | 1.11 |
| 9 | Third quartile ($Q_3$) | 1.58 | 1.40 | 1.36 | 1.38 | 1.33 | 1.31 | 1.30 | 1.28 | 1.31 |
| 10 | Highest-error forecast (MAX) | 4.84 | 3.87 | 6.45 | 5.76 | 6.69 | 3.29 | 3.77 | 4.40 | 3.84 |
| | | Implicit Price Deflator (IPD) | | | | | | | | |
| 11 | Lowest-error forecast (MIN) | .83 | .88 | .82 | .55 | .53 | .67 | .88 | .81 | .71 |
| 12 | First quartile ($Q_1$) | 1.13 | 1.03 | 1.02 | .99 | 1.00 | 1.03 | 1.02 | 1.02 | 1.02 |
| 13 | Median (MD) | 1.24 | 1.21 | 1.15 | 1.12 | 1.12 | 1.17 | 1.12 | 1.11 | 1.11 |
| 14 | Third quartile ($Q_3$) | 1.56 | 1.39 | 1.34 | 1.27 | 1.23 | 1.39 | 1.32 | 1.23 | 1.29 |
| 15 | Highest-error forecast (MAX) | 3.55 | 4.30 | 3.55 | 3.74 | 3.37 | 3.27 | 2.68 | 2.86 | 3.71 |

*Note:* All entries show ratios $RMSE_i/RMSE_g$, where the subscripts i and g refer to the individual and group mean forecasts, respectively. MIN and MAX denote the lowest and highest ratios in each distribution, $Q_1$ and $Q_3$ denote the lower- and upper-quartile ratios, and MD denotes the median ratio.

3 and 9). The RMSEs of the group mean forecasts are about 23–29 percent of the RMSVs for nominal GNP growth and inflation and 51–53 percent of the RMSVs for real GNP growth (cf. cols. 6 and 9).

### 1.3.5 Have Any Forecasters Excelled Consistently?

Each forecaster in our sample of 111 was ranked by the accuracy of his or her predictions, separately for each forecast target as defined by the date of the survey ($t$), variable, and span (e.g., for the GNP 0–1 predictions made in 1970:I). Let $r_{it}$ be the rank of the $i$th respondent in the time $t$ survey, which increases from the smallest to the largest squared error. The number of surveys covered per respondent ($m_i$) varied widely across individuals, and the number of respondents per survey ($n_t$) varied widely across time (see table 1.2, col. 6). In view of this variability, it was necessary to normalize the ranks by the number of participants in the particular survey. This is done by calculating $R_t = 100r_{it}/n_t$. The best forecast in each set would have $r_{it} = 1$, and hence $R_{it} = 100/n_t$. The worst forecast would have $r_{it} = n_t$, and hence $R_{it} = 100$. This setup permits us to consider the question, How stable were the accuracy rankings of the forecasters over time?

When the ranks are aggregated across the corresponding sets for each individual, measures of central tendency and dispersion are obtained that characterize the distributions over time of the ranks. Thus, for a given variable and span, the overall rank of the $i$th forecaster is $R_i = 1/m \, \Sigma_t^m R_{it}$, and the corresponding standard deviation equals $[1/m \, \Sigma_t^m (R_{it} - R_i)^2]^{1/2}$. We compute such means, standard deviations, medians, quartiles, and ranges for each of the 111 individuals covered. Table 1.11 presents simple averages of some of these measures in columns 1–4. For example, the grand means (Ms) in column 1 represent $\bar{R} = 1/111 \, \Sigma_i R_i$.

In addition, columns 5–10 of table 1.11 summarize the distributions across individuals of the mean normalized ranks $R_i$. The selected statistics include SD, quartiles, and extremes. For example, here $SD = [1/111 \, \Sigma_i (R_i - \bar{R})^2]^{1/2}$ (col. 5).

The entries in column 1 are all very close: 53–55 for GNP, 52 for RGNP, and 54–59 for IPD. The corresponding medians (not shown) are similarly clustered but one to two points larger. In fact, there is very little variation between the entries in any column of table 1.11. That is, the distributions of the normalized ranks are very similar for any of the three variables covered and for any of the five spans.

Typically, any forecaster would rank high at some times and low at others. Indeed, the *average* range of 85–90 (col. 4) is close to the *maximum* range possible for the $R_{it}$ ranks (which cannot exceed 99 and would not be much larger than 90 for relatively small values of $n_t$). The forecaster's rank would fall in the center half of the distribution (i.e., in the interquartile range IQR) nearly 50 percent of the time and within ± SD of the mean perhaps up to 66 percent of the time (cols. 2 and 3). There is no evidence of a high skewness or

**Table 1.11** **Ranking Forecasters According to Their Accuracy in Predicting Nominal and Real GNP and Inflation Rates: Selected Measures, 1968–90**

| | | Distribution over Time of Individual Normalized Ranks ($R_{it}$): Means of: | | | | Distribution across Individuals of Mean Normalized Ranks ($R_i$) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Row | Span (Qs) | M (1) | SD (2) | IQR (3) | Range (4) | SD (5) | MIN (6) | $Q_1$ (7) | MD (8) | $Q_3$ (9) | MAX (10) |
| | | Gross National Product (GNP) | | | | | | | | | |
| 1 | 0–1 | 53 | 27 | 46 | 87 | 10 | 38 | 48 | 53 | 59 | 95 |
| 2 | 0–2 | 53 | 27 | 46 | 87 | 11 | 32 | 45 | 51 | 59 | 98 |
| 3 | 0–3 | 53 | 27 | 46 | 88 | 11 | 36 | 46 | 51 | 58 | 100 |
| 4 | 0–4 | 53 | 27 | 46 | 86 | 11 | 33 | 47 | 52 | 58 | 99 |
| 5 | 0–5 | 53 | 26 | 45 | 85 | 12 | 30 | 46 | 51 | 58 | 100 |
| | | Gross National Product in Constant Dollars (RGNP) | | | | | | | | | |
| 6 | 0–1 | 52 | 28 | 50 | 90 | 9 | 29 | 47 | 52 | 56 | 77 |
| 7 | 0–2 | 52 | 28 | 47 | 89 | 10 | 30 | 47 | 51 | 57 | 94 |
| 8 | 0–3 | 52 | 28 | 47 | 88 | 9 | 35 | 46 | 51 | 57 | 89 |
| 9 | 0–4 | 52 | 28 | 48 | 88 | 9 | 34 | 45 | 51 | 58 | 87 |
| 10 | 0–5 | 52 | 28 | 48 | 86 | 10 | 31 | 45 | 51 | 57 | 99 |
| | | Implicit Price Deflator (IPD) | | | | | | | | | |
| 11 | 0–1 | 59 | 28 | 48 | 89 | 10 | 40 | 52 | 57 | 65 | 93 |
| 12 | 0–2 | 56 | 28 | 48 | 89 | 10 | 39 | 49 | 56 | 62 | 95 |
| 13 | 0–3 | 55 | 28 | 47 | 88 | 10 | 33 | 48 | 54 | 60 | 98 |
| 14 | 0–4 | 54 | 28 | 46 | 88 | 10 | 29 | 48 | 54 | 60 | 93 |
| 15 | 0–5 | 54 | 27 | 46 | 86 | 11 | 29 | 45 | 54 | 59 | 94 |

*Note:* The basic unit of measurement is the normalized rank $R_{it} = 100(r_{it}/n_t)$, where $r_{it}$ = rank of the $i$th forecaster in the time $t$ set of predictions for a given variable and span, and $n_t$ = number of forecasters in the same set. The ranks are assigned according to the squared errors $(P - A)^2$, from the smallest to the largest. The entries in cols. 1–4 represent the means of the summary measures for the distributions of the individuals' ranks over time (e.g., M in col. 1 refers to $\bar{R} = 1/n\ \Sigma_i^n R_i$, where $R_i = 1/m\ \Sigma_t^m R_{it}$; similarly for the standard deviations in col. 2, etc.). The entries in cols. 5–10 characterize the distributions across the individuals of $R_i$. All statistics are rounded off, with no decimals shown. For symbols, see the preceding tables and the text.

a high kurtosis in these distributions. To sum up, the forecasting performance of any one individual relative to another is likely to be highly variable over time.

On the other hand, the dispersion of the corresponding forecasts and their errors across individuals will tend to be limited by the commonality of the targets of the forecasters and of the information and methods available to them. The correlations between the forecasters' errors are expected to be positive and may be high. Our measures presumably reflect all these regularities. Interestingly, the standard deviations in column 2 are 26–28, but those in column 5 are only 9–12 (note that the definition of the former includes time $t$ explicitly, while the definition of the latter does not). Similarly, the IQRs in column 3 are 45–50, those implied by columns 7 and 9 are 9–15, and the corresponding total ranges are 85–90 and 30–59 (cf. cols. 4 and 6–10). These numbers seem consistent with the results obtained in some previous studies indicating that fluctuations over time contribute more than differences across forecasters to the overall variation in forecast errors (see Zarnowitz 1974, 578–79).

For each of the forecast targets identified in lines 1–15 of table 1.11, the ranks according to $R_i$ form a relatively tight cluster between the values of $Q_1$ and $Q_3$ that average 47 and 59, respectively (cols. 7–9). One-quarter of the group performed poorly relative to the others, with $R_i$ values ranging from well above 60 to 100 (cols. 9–10). However, our attention centers on the top-ranking quarter, with $R_i$ values averaging in the 30s and 40s (cols. 6–7). The latter can be said to have excelled with respect to the given category of forecast targets.

All these subsets, of course, consist of individuals who are coded and can be identified. It is important to ask next what the correlations of the ranks are between the different variables and spans. For example, do those who best predicted the growth of real GNP also tend to excel in predicting inflation? Do those who rank high in forecasting over the shortest horizons also rank high in forecasting over the longer horizons?

Table 1.12 indicates that the answers to these questions are on the whole positive. The correlations among our normalized ranks, both across the variables for each span (rows 1–3) and across the spans for each variable (rows 4–13), are all positive and sufficiently high not to be due to chance. Forecasters who predict relatively well (poorly) any one of these targets are also likely to predict well (poorly) any of the other targets. Not surprisingly, the correlations are higher the more closely related the forecast targets are. Thus, they are higher for GNP and RGNP than for RGNP and IPD and higher for successive spans (e.g., 0–1 and 0–2) than for more distant spans (e.g., 0–1 and 0–5). Similar results have been found for other variables and periods and for marginal as well as average change forecasts (cf. Zarnowitz 1984, 17–19).

**Table 1.12 Respondents to NBER-ASA Surveys Ranked According to the Accuracy of Their Forecasts of Nominal and Real GNP Growth and IPD Inflation Rates: Correlations among the Ranks and across Variables and Horizons, 1968–90**

| | | Correlated for Forecast Horizons (in Qs) | | | | |
|---|---|---|---|---|---|---|
| Row | Variables | 0–1 (1) | 0–2 (2) | 0–3 (3) | 0–4 (4) | 0–5 (5) |
| 1 | GNP, RGNP | .73 | .74 | .68 | .64 | .64 |
| 2 | GNP, IP | .56 | .64 | .68 | .59 | .59 |
| 3 | RGNP, IPD | .47 | .52 | .54 | .51 | .42 |
| | | Correlated for Variables | | | | |
| | Horizons | GNP | RGNP | IPD | | |
| 4 | 0–1, 0–2 | .82 | .77 | .79 | | |
| 5 | 0–1, 0–3 | .73 | .68 | .65 | | |
| 6 | 0–1, 0–4 | .68 | .67 | .51 | | |
| 7 | 0–1, 0–5 | .73 | .62 | .47 | | |
| 8 | 0–2, 0–3 | .87 | .87 | .83 | | |
| 9 | 0–2, 0–4 | .75 | .74 | .75 | | |
| 10 | 0–2, 0–5 | .79 | .72 | .67 | | |
| 11 | 0–3, 0–4 | .92 | .86 | .92 | | |
| 12 | 0–3, 0–5 | .87 | .80 | .86 | | |
| 13 | 0–4, 0–5 | .92 | .86 | .92 | | |

*Note:* The correlations are based on the normalized ranks described in the text and table 1.11. On the symbols used, see previous tables and the text.

### 1.3.6 Comparisons with Bayesian Vector Autoregressive (BVAR) Forecasts

We use a BVAR model with five variables: RGNP, IPD, M2 (broad money supply), LI (the composite index of leading indicators), and TBR (the three-month Treasury-bill rate). TBR is a level series; the others are series of growth rates. The model is estimated on quarterly series, each taken with six lags. The data are those presently available (i.e., they incorporate all revisions), and, in this sense, the forecasts based on them are ex post. But the forecasts are generated sequentially, using only the information preceding the date of the forecast.

Unlike the forecasters, who can take advantage of the early information provided by the monthly and weekly time series released during the survey quarter, the BVAR model does not draw on any such data. On the other hand, unlike the BVAR model, which is based on the present, revised series, the forecasters work under the disadvantage of having access only to the latest preliminary data, that is, data containing measurement errors that have yet to be eliminated by revisions.

Because the quarterly data for the survey quarter (1) are not known to the forecasters, our first approach was to impute the same lack of knowledge to our BVAR model. Here, then, the shortest prediction is for 0–1, the longest for 0–5. But, as pointed out by Christopher Sims during the conference, this approach (now called "variant A") ignores any effects on the survey forecasts of the most recent economic news. Since the knowledge of the news on balance presumably helps the forecasters, variant A in this respect handicaps our BVAR, as it would more generally any model based strictly on quarterly time series only.

For this reason, we also present the results of alternative calculations ("variant B"), which assume full knowledge of the actual values in quarter 1, or effectively perfect foresight. Here, for 0–1, the error of the BVAR model is identically zero, and no comparisons with the survey forecasts are available; the shortest prediction is for 1–2. Thus, the two variants represent contrasting extremes: in A there is no knowledge; in B there is full knowledge of period 1 values. Variant B handicaps the real-life forecaster, who has only partial and indirect knowledge of the target variable in the current (survey) quarter.

It follows that the truth about the relative accuracy of the individual forecasts from the surveys and the BVAR forecasts falls somewhere between variants A and B. Table 1.13 provides the evidence, showing in columns 1–3 that the measures of error of BVAR variant A for spans 0–1, 0–2, . . . , etc. are approximately equal to the corresponding measures of error of BVAR variant B for spans 0–2, 0–3, . . . , etc., respectively (cf. rows 1 and 7, 2 and 8, and so on). As would be expected, the RMSE ratios in columns 4–8 are throughout lower for variant A than for variant B, when comparing entries for the corresponding spans (rows 2 and 7, 3 and 8, and so on). That is, variant B calculations show the BVAR model forecasts in a relatively more favorable light than variant A calculations do.

We present the results for both variants of the retroactively used time-series models for comparisons relating to GNP, RGNP, and IPD (this covers both our own and outside, multivariate and univariate models). For the other variables, only variant A is used. More often than not, the "true" outcomes are probably closer to the variant A than to the variant B comparisons because the forecasters' information about recent and current developments is in fact quite limited and deficient and because the forecasters use preliminary data and the time-series models use revised data. When all is considered, it can be argued that variant B handicaps the forecasters more than variant A handicaps the models.

The RMSE ratios in table 1.13, columns 4–8, indicate that at least 75 percent of the individual forecasts of GNP, 50 percent of those of IPD, and 25 percent of those of RGNP were more accurate than the variant A BVAR forecasts. Thus, the $Q_3$ ratios are less than 1.0 for nominal growth and close to 1.0 for inflation. For real growth, the MD ratios approach unity at spans of 2 to 3 quarters and exceed it at longer spans. The ratios based on the variant B

Table 1.13 **BVAR Forecasts (Two Variants) versus Individual Forecasts from NBER-ASA Surveys: Summary Measures of Error and RMSE Ratios for GNP, RGNP, and IPD, 1968–90**

| | | BVAR Forecasts[a] | | | RMSE Ratios, Individual to BVAR Forecasts[b] | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Row | Span (Qs) | M (1) | MAE (2) | RMSE (3) | MIN (4) | $Q_1$ (5) | MD (6) | $Q_3$ (7) | MAX (8) |
| | | Gross National Product (GNP): Variant A | | | | | | | |
| 1 | 0–1 | .07 | .84 | 1.11 | .32 | .54 | .66 | .91 | 1.89 |
| 2 | 0–2 | .18 | 1.47 | 1.92 | .34 | .57 | .68 | .83 | 2.00 |
| 3 | 0–3 | .26 | 2.08 | 2.73 | .38 | .58 | .70 | .83 | 2.44 |
| 4 | 0–4 | .33 | 2.59 | 3.45 | .24 | .59 | .68 | .85 | 2.61 |
| 5 | 0–5 | .38 | 3.23 | 4.23 | .25 | .54 | .67 | .82 | 2.41 |
| | | GNP: Variant B | | | | | | | |
| 6 | 0–1 | 0 | 0 | 0 | N.A. | N.A. | N.A. | N.A. | N.A. |
| 7 | 0–2 | .07 | .87 | 1.14 | .43 | .75 | .85 | 1.03 | 2.89 |
| 8 | 0–3 | .17 | 1.49 | 1.96 | .53 | .78 | .91 | 1.03 | 1.90 |
| 9 | 0–4 | .26 | 2.12 | 2.80 | .49 | .81 | .91 | 1.05 | 1.82 |
| 10 | 0–5 | .33 | 2.63 | 3.53 | .63 | .86 | .96 | 1.13 | 2.29 |
| | | GNP in Constant Dollars (RGNP): Variant A | | | | | | | |
| 11 | 0–1 | .08 | .78 | 1.00 | .29 | .59 | .75 | .99 | 3.65 |
| 12 | 0–2 | .20 | 1.09 | 1.51 | .36 | .73 | .89 | 1.07 | 2.25 |
| 13 | 0–3 | .28 | 1.53 | 2.03 | .43 | .79 | .93 | 1.13 | 2.36 |
| 14 | 0–4 | .35 | 1.76 | 2.34 | .52 | .96 | 1.08 | 1.29 | 2.74 |
| 15 | 0–5 | .39 | 2.05 | 2.64 | .41 | 1.00 | 1.13 | 1.40 | 2.90 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | RGNP: Variant B | | | | | | | |
| 16 | 0–1 | 0 | 0 | 0 | N.A. | N.A. | N.A. | N.A. | N.A. |
| 17 | 0–2 | .09 | .78 | 1.01 | .48 | .83 | .96 | 1.16 | 3.87 |
| 18 | 0–3 | .20 | 1.11 | 1.53 | .46 | 1.00 | 1.15 | 1.33 | 2.85 |
| 19 | 0–4 | .29 | 1.56 | 2.06 | .63 | 1.06 | 1.19 | 1.37 | 3.68 |
| 20 | 0–5 | .36 | 1.79 | 2.38 | .71 | 1.18 | 1.40 | 1.61 | 3.14 |
| | | Implicit Price Deflator (IPD): Variant A | | | | | | | |
| 21 | 0–1 | .05 | .37 | .48 | .55 | .81 | .97 | 1.16 | 3.68 |
| 22 | 0–2 | .11 | .76 | .97 | .49 | .76 | .87 | 1.02 | 3.62 |
| 23 | 0–3 | .17 | 1.18 | 1.53 | .40 | .72 | .86 | 1.02 | 2.95 |
| 24 | 0–4 | .23 | 1.65 | 2.18 | .38 | .72 | .87 | 1.04 | 2.74 |
| 25 | 0–5 | .28 | 2.19 | 2.94 | .37 | .72 | .86 | 1.06 | 3.94 |
| | | IPD: Variant B | | | | | | | |
| 26 | 0–1 | 0 | 0 | 0 | N.A. | N.A. | N.A. | N.A. | N.A. |
| 27 | 0–2 | .04 | .37 | .47 | .70 | 1.07 | 1.23 | 1.51 | 3.69 |
| 28 | 0–3 | .10 | .76 | .97 | .71 | .95 | 1.09 | 1.30 | 2.83 |
| 29 | 0–4 | .16 | 1.17 | 1.53 | .57 | .98 | 1.10 | 1.24 | 2.56 |
| 30 | 0–5 | .21 | 1.64 | 2.18 | .61 | .95 | 1.08 | 1.24 | 4.23 |

*Note:* N.A. = not available.

[a]Based on a model with five variables (RGNP, IPD, M2, LI, and TBR) and six quarterly lags, estimated sequentially with presently available data. Variant A assumes that the last known values of the variables to be predicted refer to the quarter $t - 1$ (denoted 0); variant B assumes that they refer to the current quarter $t$ (denoted 1).

[b]$RMSE_i/RMSE_{bv}$, where the subscript i refers to the individual forecasts from the NBER-ASA surveys and the subscript bv refers to the corresponding Bayesian vector autoregressive (BVAR) forecasts (variant A in rows 1–5, 11–15, and 21–25; variant B in rows 6–10, 16–20, and 26–30). MIN and MAX denote the lowest and highest ratios in each distribution, $Q_1$ and $Q_3$ denote the lower- and upper-quartile ratios, and MD denotes the median ratio.

BVAR forecasts still show most of the survey forecasts to be superior for GNP, but not for IPD or RGNP. Here the ratios rise above 1.0 for all horizons at $Q_3$ for GNP, at MD for IPD, and even at $Q_1$ for RGNP.

The BVAR mean errors are all positive, unlike the MEs for the NBER-ASA survey forecasts, which are mostly negative for GNP and IPD and mostly positive but somewhat mixed for RGNP. (For this and the rest of the paragraph, see table 1.13, cols. 1–3, and table 1.9, cols. 1–6.) Comparisons of the MAEs and RMSEs of BVAR with the corresponding measures for the average individual survey forecast produce a mixed picture, depending on the series and criteria used. However, the comparisons with the group means are generally adverse for BVAR of either variant.

Such variables as the leading index and the short-term interest rate act as strong codeterminants of growth in total output, as suggested by regression estimates and out-of-sample predictions with VAR models (Zarnowitz and Braun 1990; Zarnowitz 1992, chap. 11). Our findings here are consistent with these results. The BVAR forecasts of RGNP perform relatively well, which holds a potentially useful lesson for the forecasters to take proper account of these relations. But the BVAR forecasts of GNP and IPD are apparently weaker.

### 1.3.7 Comparing Forecasts for the First and Second Halves of 1968–90

The period 1968:IV–1979:III was one of upward drifts and large instability in both inflation and unemployment; of business contractions in 1969–70 and 1973–75; of the Vietnam War and price control disturbances in the early years; and of severe supply (mainly oil price) shocks in the middle and late years. The period 1979:IV–1990:I was one of more successful attempts to slow inflation by restrictive monetary policy; of sharp rises in prices and interest rates followed by downward trends in the wake of two back-to-back recessions in 1980 and 1981–82; of a long expansion that followed, interrupted by slowdowns in 1984–86 and 1989; and of new trade and financial problems. It is of interest to ask how the macro forecasts fared in these two so different periods of approximately equal length.

The errors of the individual forecasts from the NBER-ASA surveys were on average larger in 1979–90 than in 1968–79 for GNP but smaller for IPD, judging from the comparisons of the RMSEs in table 1.14, columns 1 and 5. For RGNP, the differences between the two subperiods are small and mixed, depending on the horizon of the forecasts.

The average individual (i)-to-group mean (g) RMSE ratios differ little between 1968–79 ($1.04 \leq i/g \geq 1.34$) and 1979–90 ($1.15 \leq i/g \geq 1.31$). They decreased somewhat in the latter period for short GNP and RGNP forecasts, increased more for longer IPD forecasts, but remained approximately unchanged in most cases (cf. cols. 2 and 6).

The individual-to-BVAR (bv) RMSE ratios for GNP rose from 0.6 or less in 1968–79 to around 0.8 in 1979–90; those for RGNP rose as well, from an

Table 1.14 **Individual, Group Mean, and BVAR Forecasts of Percentage Changes in GNP, RGNP, and IPD: Selected Comparisons by Span and Subperiods, 1968–79 and 1979–90**

| | | Forecasts for 1968:IV–1979:III | | | | Forecasts for 1979:IV–1990:I | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Median | RMSE Ratios | | | Median | RMSE Ratios[b] | | |
| Rows | Span (Qs) | $RMSE_i$[a] (1) | i/g (2) | i/bv (3) | g/bv (4) | $RMSE_i$[a] (5) | i/g (6) | i/bv (7) | g/bv (8) |
| | | Gross National Product (GNP) | | | | | | | |
| 1 | 0–1 | .60 | 1.34 | .53 | .48 | .86 | 1.15 | .79 | .71 |
| 2 | 0–2 | 1.13 | 1.20 | .58 | .50 | 1.56 | 1.17 | .78 | .69 |
| 3 | 0–3 | 1.68 | 1.18 | .60 | .51 | 2.23 | 1.18 | .81 | .70 |
| 4 | 0–4 | 2.04 | 1.21 | .60 | .49 | 3.08 | 1.15 | .81 | .71 |
| 5 | 0–5 | 2.24 | 1.19 | .51 | .47 | 3.80 | 1.18 | .82 | .70 |
| | | Gross National Product in Constant Dollars (RGNP) | | | | | | | |
| 6 | 0–1 | .69 | 1.29 | .62 | .57 | .80 | 1.20 | .88 | .76 |
| 7 | 0–2 | 1.25 | 1.18 | .78 | .69 | 1.34 | 1.15 | 1.01 | .82 |
| 8 | 0–3 | 1.87 | 1.17 | .85 | .76 | 1.80 | 1.15 | 1.03 | .86 |
| 9 | 0–4 | 2.59 | 1.14 | .99 | .84 | 2.30 | 1.16 | 1.16 | .94 |
| 10 | 0–5 | 3.13 | 1.13 | 1.04 | .92 | 2.84 | 1.15 | 1.24 | .99 |
| | | Implicit Price Deflator (IPD) | | | | | | | |
| 11 | 0–1 | .50 | 1.29 | .90 | .73 | .37 | 1.22 | .87 | .74 |
| 12 | 0–2 | 1.00 | 1.14 | .90 | .80 | .64 | 1.31 | .70 | .56 |
| 13 | 0–3 | 1.57 | 1.08 | .89 | .84 | .93 | 1.28 | .60 | .51 |
| 14 | 0–4 | 2.25 | 1.04 | .92 | .89 | 1.37 | 1.26 | .64 | .50 |
| 15 | 0–5 | 3.06 | 1.05 | .99 | .95 | 1.94 | 1.25 | .63 | .49 |

[a]Median of the root mean square errors of the individual forecasts from the quarterly NBER-ASA surveys.

[b]Ratio of the median RMSE of the individual forecasts (i) to the RMSE of the corresponding group mean forecast (g) in cols. 2 and 6. Ratio of the median RMSE of the individual forecasts (i) to the RMSE of the corresponding BVAR model forecast (bv) in cols. 3 and 7. Ratio of the RMSE of the group mean forecast (g) to the RMSE of the corresponding BVAR model (bv) in cols. 4 and 8.

approximate range of 0.6–1.0 to 0.9–1.2; and those for IPD declined from 0.9–1.0 to 0.6–0.9 (cols. 3 and 7). These i/bv ratios, then, show that, on average, the NBER-ASA survey forecasts outperformed our BVAR forecasts, except for RGNP in 1979–90. The group mean predictions from the surveys were throughout more accurate than BVAR; that is, the ratios g/bv are less than one in all cases (cols. 4 and 8). As might be expected, the changes in i/bv and g/bv between the two subperiods paralleled each other directionally.

There is no evidence here that, on the whole, the forecasts either improved or deteriorated in the 1980s as compared with the 1970s. The BVAR benchmark proved a little more effective in 1979–90 than in 1968–79 for nominal and real GNP growth and somewhat less effective for inflation.

## 1.4 Other Forecasts for 1968–90

### 1.4.1 Percentage Change Forecasts: Industrial Production and Corporate Profits

Table 1.15 shows that the average errors of the forecasts of IP (industrial production) and CP (corporate profits) tended to be positive but widely dispersed and strongly increasing with the span (cols. 1–3). The RMSEs increased in a similar fashion (cols. 4–6). Comparisons with the average size and variability of the actual changes (cols. 9–11) indicate a moderate level of accuracy for the IP forecasts but poor overall performance for the CP forecasts (where the mean and median RMSEs exceed the actual SD and RMSV values). The large positive values of SK and KU for the IP predictions up to three quarters ahead suggest skewness to the right and fat tails; the latter may also characterize the longer CP predictions (cols. 7–8).

Combining the individual forecasts by simple averaging reduces the errors substantially for IP (except for the longest span) but not for CP, where the gains from using the group mean or consensus forecast are small (cf. table 1.15, cols. 4 and 6, with table 1.16, col. 1). Accordingly, the RMSE ratios i/g are smaller for CP than for IP, but it is still true for both variables that only about the best 25 percent of the sample are more accurate than the group mean forecasts (see table 1.16, cols. 2–4).

The BVAR model forecasts (variant A only) outperform the group mean forecasts for profits. The comparisons for the production index yield closer and mixed results, which favor the survey group's predictions for the shorter and the BVAR predictions for the longer horizons. (Compare the corresponding entries in cols. 1–4 and 5–8 of table 1.16.)

In almost all cases, both IP and CP forecasts had larger RMSEs in 1979–90 than in 1968–79 (table 1.17, cols. 1 and 5). Compared with BVAR variant A, the survey forecasts look better in the earlier than in the later subperiod, particularly for IP (cf. cols. 3 and 4 with cols. 7 and 8, respectively).

**Table 1.15** **Selected Measures of Forecast Accuracy and Actual Values: Percentage Changes in Industrial Production and Corporate Profits, by Span, 1968–90**

| | | Mean Error | | | Root Mean Square Error | | | | | Actual Value[a] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Row | Span (Qs) | M (1) | SD (2) | MD (3) | M (4) | SD (5) | MD (6) | SK (7) | KU (8) | M (9) | SD (10) | RMSV (11) |
| | | | | | Index of Industrial Production (IP)[b] | | | | | | | |
| 1 | 0–1 | .04 | .58 | .02 | 1.66 | 1.08 | 1.54 | 7.89 | 73.16 | .76 | 2.17 | 2.29 |
| 2 | 0–2 | .83 | 1.07 | .30 | 3.13 | 1.80 | 2.93 | 7.66 | 69.68 | 1.52 | 3.79 | 4.08 |
| 3 | 0–3 | .67 | 1.58 | .63 | 4.52 | 2.10 | 4.26 | 6.28 | 52.71 | 2.28 | 5.04 | 5.53 |
| 4 | 0–4 | 1.04 | 1.90 | 1.09 | 5.45 | 1.35 | 5.34 | .80 | 1.58 | 3.06 | 6.08 | 6.81 |
| 5 | 0–5 | 1.06 | 2.27 | 1.35 | 6.19 | 1.37 | 6.02 | .36 | 1.05 | 3.83 | 6.95 | 7.94 |
| | | | | | Corporate Profits after Taxes (CP)[c] | | | | | | | |
| 6 | 0–1 | .26 | 2.49 | −.02 | 9.50 | 2.03 | 9.39 | .04 | 1.08 | 1.33 | 7.36 | 7.48 |
| 7 | 0–2 | 1.00 | 4.58 | .76 | 14.42 | 2.86 | 14.71 | −.39 | 1.95 | 2.78 | 11.13 | 11.47 |
| 8 | 0–3 | 2.58 | 6.54 | 2.64 | 18.58 | 3.32 | 18.75 | −.14 | 4.23 | 4.17 | 13.54 | 14.17 |
| 9 | 0–4 | 4.11 | 8.20 | 4.39 | 22.38 | 4.29 | 22.58 | .05 | 7.34 | 5.55 | 15.56 | 16.52 |
| 10 | 0–5 | 6.29 | 9.66 | 6.41 | 26.30 | 4.94 | 26.47 | .93 | 9.07 | 7.01 | 17.82 | 19.15 |

[a]Refers to the period 1970:I–1989:IV.

[b]Based on the second revision of the monthly data.

[c]Based on the first July revision of the quarterly data.

Table 1.16 **Individual, Group Mean, and BVAR Forecasts of Percentage Changes in Industrial Production and Corporate Profits: Selected Comparisons, by Span, 1968–90**

| Row | Span (Qs) | Group Mean RMSE (1) | RMSE Ratios i/g $Q_1$ (2) | RMSE Ratios i/g MD (3) | RMSE Ratios i/g $Q_3$ (4) | BVAR[a] RMSE (5) | RMSE Ratios i/bv $Q_1$ (6) | RMSE Ratios i/bv MD (7) | RMSE Ratios i/bv $Q_3$ (8) |
|---|---|---|---|---|---|---|---|---|---|
| | | Index of Industrial Production (IP)[b] | | | | | | | |
| 1 | 0–1 | 1.17 | 1.13 | 1.26 | 1.54 | 1.56 | .85 | 1.00 | 1.32 |
| 2 | 0–2 | 2.44 | 1.06 | 1.16 | 1.30 | 2.83 | .90 | 1.04 | 1.45 |
| 3 | 0–3 | 3.50 | 1.07 | 1.16 | 1.28 | 3.66 | .99 | 1.20 | 1.58 |
| 4 | 0–4 | 4.55 | 1.06 | 1.13 | 1.24 | 4.25 | 1.04 | 1.25 | 1.74 |
| 5 | 0–5 | 6.16 | 1.05 | 1.13 | 1.26 | 4.78 | 1.04 | 1.21 | 1.66 |
| | | Corporate Profits after Taxes (CIP)[c] | | | | | | | |
| 6 | 0–1 | 9.24 | 1.00 | 1.08 | 1.14 | 7.22 | 1.25 | 1.41 | 1.58 |
| 7 | 0–2 | 13.68 | 1.01 | 1.08 | 1.16 | 11.22 | 1.15 | 1.32 | 1.49 |
| 8 | 0–3 | 17.36 | 1.00 | 1.06 | 1.14 | 14.35 | 1.14 | 1.29 | 1.44 |
| 9 | 0–4 | 20.98 | 1.00 | 1.06 | 1.14 | 16.62 | 1.18 | 1.31 | 1.48 |
| 10 | 0–5 | 24.41 | .98 | 1.05 | 1.12 | 19.16 | 1.14 | 1.33 | 1.51 |

[a]For IP, based on a model with six variables (RGNP, IPD, M2, LI, TBR, and IP) and six quarterly lags, estimated sequentially with presently available data. For CP, based on a model with six variables (RGNP, IPD, M2, LI, TBR, and CP) and six quarterly lags, estimated sequentially with presently available data. BVAR variant A is used throughout.

[b]Based on the second revision of the monthly data.

[c]Based on the first July revision of the quarterly data.

### 1.4.2 Level Forecasts: Unemployment Rate and Housing Starts

For UR (unemployment rate; table 1.18, rows 1–5), the mean errors are predominantly negative, suggesting some underprediction, but they also show considerable dispersion. Level errors, unlike average change errors, do not cumulate, but the RMSEs still increase substantially with the distance to the target quarter. The summary error measures are quite small relative to the statistics for the actual values of UR. For short forecasts, the distributions of the RMSEs are skewed to the right and have fat tails, judging from the large SK and KU values.

For HS (housing starts; rows 6–10), the mean errors are close to zero and have mixed signs. They do not depend on the distance to the target (unlike the mean errors for UR, which increase with the distance). The RMSE and SD values, as usual, increase for the longer forecasts, but they remain fairly small compared with the measures for the actual values of HS. The SK and KU figures are small.

Combining the individual forecasts results in substantial gains in accuracy for both variables, but particularly for UR (cf. table 1.19, col. 1, and table

**Table 1.17 Individual, Group Mean, and BVAR Forecasts of Percentage Changes in Industrial Production and Corporate Profits: Selected Comparisons, by Span and Subperiod, 1968–79 and 1979–90**

| | | Forecasts for 1968:IV–1979:III | | | | Forecasts for 1979:IV–1990:I | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | RMSE Ratios | | | | RMSE Ratios | | |
| Raw | Span (Qs) | Median RMSE (1) | i/g (2) | i/bv (3) | g/bv (4) | Median RMSE (5) | i/g (6) | i/bv (7) | g/bv (8) |
| | | Index of Industrial Production (IP) | | | | | | | |
| 1 | 0–1 | 1.65 | 1.23 | 1.00 | .91 | 1.49 | 1.30 | 1.22 | .83 |
| 2 | 0–2 | 2.95 | 1.16 | 1.01 | .92 | 3.13 | 1.09 | 1.42 | 1.07 |
| 3 | 0–3 | 4.17 | 1.17 | 1.09 | .95 | 4.54 | 1.11 | 1.65 | 1.21 |
| 4 | 0–4 | 4.96 | 1.14 | 1.14 | 1.00 | 5.84 | 1.10 | 1.70 | 1.29 |
| 5 | 0–5 | 5.37 | 1.18 | 1.03 | 1.28 | 6.98 | 1.08 | 1.64 | 1.35 |
| | | Corporate Profits after Taxes (CP) | | | | | | | |
| 6 | 0–1 | 9.13 | 1.06 | 1.48 | 1.40 | 10.08 | 1.06 | 1.37 | 1.20 |
| 7 | 0–2 | 14.12 | 1.08 | 1.32 | 1.25 | 15.06 | 1.05 | 1.44 | 1.25 |
| 8 | 0–3 | 17.62 | 1.05 | 1.17 | 1.13 | 19.06 | 1.08 | 1.47 | 1.29 |
| 9 | 0–4 | 20.93 | 1.06 | 1.17 | 1.09 | 22.66 | 1.07 | 1.52 | 1.32 |
| 10 | 0–5 | 23.52 | 1.06 | 1.12 | 1.06 | 25.93 | 1.06 | 1.47 | 1.31 |

*Note:* The symbols i, g, and bv refer to the individual, group mean, and BVAR forecasts, variant A, respectively. $RMSE_i$ is the median of the RMSEs of the sampled forecasts (cols. 1 and 5). The i/g ratio is $RMSE_i/RMSE_g$ for strictly matching observations, and the i/bv ratio is $RMSE_i/RMSE_{bv}$, with medians of the individual forecasts used in each case (cols. 2 and 6 and cols. 3 and 7, respectively). The g/bv ratio is $RMSE_g/RMSE_{bv}$ (cols. 4 and 8). See also the notes to tables 1.13 and 1.14.

1.18, cols. 4 and 6). The RMSE ratios i/g are generally higher for UR than for HS, but, once again, the $Q_1$ ratios are close to one throughout; that is, about 75 percent of the individual forecasts are less accurate than the group means in either case (table 1.19, cols. 2–4). The variant A BVAR forecasts are about as accurate as the group mean forecasts for target quarters 3–5 of both UR and HS; for closer targets, the comparisons favor the surveys for UR and the BVAR for HS (cf. the corresponding entries in cols. 1-4 and 5–8).

Table 1.20 shows that, on the whole, the NBER-ASA forecasters predicted UR somewhat better and HS somewhat worse in 1968–79 than in 1979–90 (cf. cols. 1 and 5). The relative performance of the group mean and the individual forecasts was very similar in the two periods (cols. 2 and 6); that of the BVAR variant A model improved in most cases for UR but showed no systematic change for HS (cols. 3–7 and 4–8).

Table 1.18 **Selected Measures of Forecast Accuracy and Actual Values: Levels of the Unemployment Rate and Housing Starts, by Target Quarter, 1968–90**

| | | Mean Error | | | Root Mean Square Error | | | | | Actual Value | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Row | Target Quarter | M (1) | SD (2) | MD (3) | M (4) | SD (5) | MD (6) | SK (7) | KU (8) | M (9) | SD (10) | RMSV (11) |
| | | Unemployment Rate (UR)[a] | | | | | | | | | | |
| 1 | 1 | .02 | .08 | .03 | .26 | .21 | .21 | 4.93 | 27.12 | 6.50 | 1.61 | 6.70 |
| 2 | 2 | −.01 | .13 | .01 | .52 | .20 | .49 | 3.83 | 19.92 | | | |
| 3 | 3 | −.08 | .20 | −.07 | .77 | .23 | .73 | 2.31 | 11.73 | | | |
| 4 | 4 | −.20 | .28 | −.19 | .98 | .26 | .97 | 1.22 | 5.93 | | | |
| 5 | 5 | −.22 | .34 | −.29 | 1.15 | .25 | 1.13 | .53 | .65 | | | |
| | | Housing Starts (HS)[b] | | | | | | | | | | |
| 6 | 1 | −.04 | .06 | −.03 | .23 | .04 | .23 | −.01 | .08 | 1.65 | .38 | 1.69 |
| 7 | 2 | −.03 | .09 | −.02 | .29 | .05 | .29 | .18 | .67 | | | |
| 8 | 3 | −.00 | .12 | .01 | .34 | .07 | .34 | .61 | 2.23 | | | |
| 9 | 4 | .03 | .15 | .03 | .38 | .09 | .38 | .72 | 3.32 | | | |
| 10 | 5 | .06 | .18 | .08 | .42 | .10 | .41 | .69 | 2.45 | | | |

[a]Based on presently available data (no important revisions).

[b]Based on the second revision of the monthly data.

**Table 1.19 Individual, Group Mean, and BVAR Forecasts of Levels of the Unemployment Rate and Housing Starts: Selected Comparisons, by Target Quarter, 1968–90**

| Row | Target Quarter | Group Mean RMSE (1) | RMSE Ratios i/g $Q_1$ (2) | RMSE Ratios i/g MD (3) | RMSE Ratios i/g $Q_3$ (4) | BVAR[a] RMSE (5) | RMSE Ratios i/bv $Q_1$ (6) | RMSE Ratios i/bv MD (7) | RMSE Ratios i/bv $Q_3$ (8) |
|---|---|---|---|---|---|---|---|---|---|
| | | Unemployment Rate (UR)[b] | | | | | | | |
| 1 | 1 | .16 | 1.19 | 1.42 | 1.73 | .28 | .72 | .84 | .96 |
| 2 | 2 | .41 | 1.05 | 1.17 | 1.32 | .50 | .89 | 1.02 | 1.23 |
| 3 | 3 | .65 | 1.01 | 1.10 | 1.23 | .66 | .96 | 1.16 | 1.38 |
| 4 | 4 | .86 | .98 | 1.09 | 1.20 | .78 | 1.07 | 1.25 | 1.50 |
| 5 | 5 | 1.00 | .99 | 1.10 | 1.20 | .85 | 1.14 | 1.30 | 1.71 |
| | | Housing Starts (HS)[c] | | | | | | | |
| 6 | 1 | .21 | .99 | 1.06 | 1.16 | .13 | 1.52 | 1.78 | 1.99 |
| 7 | 2 | .25 | 1.00 | 1.07 | 1.18 | .20 | 1.30 | 1.41 | 1.60 |
| 8 | 3 | .29 | 1.02 | 1.08 | 1.19 | .27 | 1.15 | 1.25 | 1.38 |
| 9 | 4 | .33 | .98 | 1.10 | 1.16 | .32 | .99 | 1.13 | 1.22 |
| 10 | 5 | .36 | .99 | 1.07 | 1.18 | .37 | .93 | 1.04 | 1.16 |

[a]For IP, based on a model with six variables (RGNP, IPD, M2, LI, TBR, and UR) and six quarterly lags, estimated sequentially with presently available data. For HS, based on a model with six variables (RGNP, IPD, M2, LI, TBR, and HS) and six quarterly lags, estimated sequentially with presently available data. See the text and the appendix. BVAR variant A is used throughout.

[b]Based on presently available data.

[c]Based on the second revision of the monthly data.

## 1.5 Comparisons with Selected Econometric and Time-Series Model Forecasts

### 1.5.1 The University of Michigan Research Seminar in Quantitative Economics

The Michigan Research Seminar in Quantitative Economics (RSQE) has the longest record of the several well-known service bureaus working with macroeconometric forecasting models. RSQE kindly provided us with the record of their forecasts, and we were able to compare them with the NBER-ASA survey forecasts for ten variables. It is important to note that the quarterly Michigan forecasts begin in 1970:IV and were not made in the first quarter in the years 1975 and 1976 and in the second quarter in the years 1971–75 and 1977–79.[12] We matched the Michigan (Mi) and the NBER-ASA (i) forecasts period by period. Further, the Michigan predictions were made typically in March, June (occasionally May), August (rarely September) and November

12. RSQE predicts normally eight times in each year.

Table 1.20 **Individual, Group Mean, and BVAR Forecasts of the Unemployment Rate and Housing Starts: Selected Comparisons, by Target Quarter and Subperiod, 1968–79 and 1979–90**

| | | Forecasts for 1968:IV–1979:III | | | | Forecasts for 1979:IV–1990:I | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | RMSE Ratios | | | | RMSE Ratios | | |
| Row | Target Quarter | Median $RMSE_i$ (1) | i/g (2) | i/bv (3) | g/bv (4) | Median $RMSE_i$ (5) | i/g (6) | i/bv (7) | g/bv (8) |
| | | Unemployment Rate (UR) | | | | | | | |
| 1 | 1 | .21 | 1.34 | .84 | .58 | .21 | 1.39 | .82 | .57 |
| 2 | 2 | .45 | 1.15 | .95 | .80 | .52 | 1.19 | 1.16 | .84 |
| 3 | 3 | .66 | 1.06 | .98 | .95 | .83 | 1.10 | 1.39 | 1.30 |
| 4 | 4 | .84 | 1.06 | 1.09 | 1.04 | 1.09 | 1.10 | 1.51 | 1.20 |
| 5 | 5 | .96 | 1.08 | 1.11 | 1.02 | 1.28 | 1.09 | 1.17 | 1.37 |
| | | Housing Starts (HS) | | | | | | | |
| 6 | 1 | .23 | 1.07 | 1.58 | 1.39 | .23 | 1.03 | 1.93 | 1.80 |
| 7 | 2 | .30 | 1.08 | 1.39 | 1.24 | .28 | 1.05 | 1.43 | 1.26 |
| 8 | 3 | .37 | 1.10 | 1.27 | 1.14 | .30 | 1.07 | 1.23 | 1.01 |
| 9 | 4 | .42 | 1.08 | 1.11 | 1.07 | .31 | 1.05 | 1.14 | .98 |
| 10 | 5 | .45 | 1.05 | 1.01 | 1.01 | .34 | 1.04 | 1.08 | .98 |

*Note:* See table 1.17.

(in 1974–75, December). The NBER-ASA survey questionnaire was usually mailed in the first half of each quarter, but it was only in the last month of the quarter that all responses were collected. Thus, at least some of the survey forecasts had the advantage of later timing (which means more potentially useful up-to-date information) vis-à-vis the Michigan forecasts.

Comparing the ME, MAE, and RMSE statistics for the Michigan and the NBER-ASA group means forecasts show the latter to have been more accurate for GNP, RGNP, and IPD (cf. cols. 1–3 and 4–6 of table 1.21). Consistent evidence comes from the RMSE ratios that have ranges of approximately 0.7–0.9, 0.9–1.1, and 1.0–1.3 for $Q_1$, MD, and $Q_3$, respectively (cols. 7–9). Thus, generally about half or more of the individual forecasts from the surveys were at least somewhat more accurate than the Michigan forecasts.

The results for the other variables are mixed. As shown in table 1.22, the Michigan predictions of real consumption show on the whole larger errors than the NBER-ASA "consensus," but not by much, and not for the longest horizon (rows 1–5). They are better than 50 percent of the individual survey forecasts for the two shortest spans and better than 75 percent for the three longest spans. The comparisons for real nonresidential investment favor the group averages by modest margins, except again for the longest span covered. For real residential investment, the Michigan forecasts are definitely better

Table 1.21 **Michigan (RSQE) Econometric Forecasts and NBER-ASA Survey Forecasts of Nominal and Real GNP Growth and IPD Inflation Rates, by Span, 1970–90**

| | | Michigan Forecasts | | | Group Mean Forecasts | | | RMSE Ratios i/Mi | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Row | Span (Qs) | M (1) | MAE (2) | RMSE (3) | M (4) | MAE (5) | RMSE (6) | $Q_1$ (7) | MD (8) | $Q_3$ (9) |
| | | Gross National Product (GNP) | | | | | | | | |
| 1 | 0–1 | −.09 | .80 | 1.08 | −.09 | .51 | .66 | .56 | .73 | .95 |
| 2 | 0–2 | .13 | 1.24 | 1.60 | −.07 | .91 | 1.18 | .73 | .89 | 1.05 |
| 3 | 0–3 | .34 | 1.45 | 1.91 | .02 | 1.33 | 1.73 | .95 | 1.11 | 1.27 |
| 4 | 0–4 | .51 | 1.81 | 2.38 | .04 | 1.64 | 2.15 | N.A. | 1.00 | 1.26 |
| 5 | 0–5 | .97 | 2.15 | 2.95 | −.02 | 1.99 | 2.61 | .76 | 1.00 | 1.19 |
| | | Gross National Product in Constant Dollars (RGNP) | | | | | | | | |
| 6 | 0–1 | .01 | .77 | 1.02 | −.05 | .51 | .66 | .56 | .75 | .99 |
| 7 | 0–2 | .25 | 1.09 | 1.49 | .01 | .88 | 1.16 | .75 | .91 | 1.13 |
| 8 | 0–3 | .46 | 1.34 | 1.77 | .14 | 1.19 | 1.64 | .91 | 1.09 | 1.28 |
| 9 | 0–4 | .77 | 1.58 | 2.18 | .19 | 1.32 | 1.89 | .81 | .97 | 1.30 |
| 10 | 0–5 | 1.20 | 1.96 | 2.88 | .29 | 1.61 | 2.32 | .75 | .94 | 1.18 |
| | | Implicit Price Deflator (IPD) | | | | | | | | |
| 11 | 0–1 | −.10 | .39 | .51 | −.03 | .27 | .34 | .71 | .88 | 1.17 |
| 12 | 0–2 | −.14 | .72 | .87 | −.08 | .52 | .68 | .81 | .97 | 1.17 |
| 13 | 0–4 | −.15 | 1.00 | 1.32 | −.12 | .76 | 1.05 | .78 | .90 | 1.05 |
| 14 | 0–4 | −.27 | 1.40 | 1.98 | −.14 | 1.12 | 1.56 | .78 | .89 | 1.02 |
| 15 | 0–5 | −.28 | 1.78 | 2.42 | −.18 | 1.57 | 2.20 | .87 | .98 | 1.12 |

*Note:* The Michigan forecasts cover the period 1970:IV–1990:I, except for the following quarters: 1971:II, 1972:II, 1973:II, 1974:II, 1975:I, 1975:II, 1976:I, 1977:II, 1978:II, and 1979:II. We match the NBER-ASA forecasts to the Michigan forecasts period by period. The ratios in cols. 7–9 are $RMSE_i/RMSE_{Mi}$, where the subscript i refers to individual forecasts from the NBER-ASA surveys and the subscript Mi refers to the Michigan forecasts.

than all but the shortest group mean forecasts. National defense expenditures are predicted better by the surveys through span 0–3 and better by Michigan (Mi) for the two longer spans. More than half of the RMSE ratios i/Mi for NFI (nonresidential fixed investment), RFI (residential fixed investment), and DEF (national defense expenditures) are less than one (rows 6–20).

The pattern that the NBER-ASA group mean forecasts have an edge for the two shortest spans and the Michigan forecasts for the two longest spans holds for the unemployment rate and the Treasury-bill rate (TBR) in table 1.23 (rows 1–5 and 6–10). The middle span shows about equal RMSEs for the two sets. The corporate bond yield (CBY) predictions from Michigan outperform those from the surveys for all but the shortest span (rows 11–15).

Table 1.22 **Michigan (RSQE) Econometric Forecasts and NBER-ASA Survey Forecasts of Percentage Changes in Consumption, Investment, and Defense Expenditures, by Span, 1981–90 and 1968–81**

| | | Michigan Forecasts | | | Group Mean Forecasts | | | RMSE Ratios i/Mi | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Row | Span (Qs) | M (1) | MAE (2) | RMSE (3) | M (4) | MAE (5) | RMSE (6) | $Q_1$ (7) | MD (8) | $Q_3$ (9) |
| | | Personal Consumption Expenditures (PCE) | | | | | | | | |
| 1 | 0–1 | −.12 | .56 | .76 | −.14 | .47 | .59 | .82 | .89 | 1.56 |
| 2 | 0–2 | −.19 | .73 | .89 | −.24 | .64 | .77 | .78 | .97 | 1.25 |
| 3 | 0–3 | −.26 | .93 | 1.15 | −.39 | .84 | .99 | .98 | 1.15 | 1.41 |
| 4 | 0–4 | −.35 | 1.10 | 1.34 | −.51 | 1.04 | 1.25 | .94 | 1.21 | 1.52 |
| 5 | 0–5 | −.41 | 1.21 | 1.51 | −.66 | 1.28 | 1.56 | .92 | 1.30 | 1.51 |
| | | Nonresidential Fixed Investment (NFI) | | | | | | | | |
| 6 | 0–1 | −.63 | 2.04 | 2.65 | −.49 | 1.68 | 2.10 | .71 | .93 | 1.18 |
| 7 | 0–2 | −1.04 | 3.25 | 4.26 | −.93 | 2.74 | 3.52 | .81 | 1.00 | 1.29 |
| 8 | 0–3 | −1.09 | 4.91 | 5.94 | −1.38 | 4.03 | 5.23 | .70 | .87 | 1.09 |
| 9 | 0–4 | −.95 | 6.47 | 7.48 | −1.71 | 5.57 | 7.09 | .76 | .93 | 1.16 |
| 10 | 0–5 | −.84 | 7.71 | 8.68 | −2.16 | 7.57 | 9.11 | .82 | 1.10 | 1.33 |
| | | Residential Fixed Investment (RFI) | | | | | | | | |
| 11 | 0–1 | −.34 | 2.53 | 3.54 | −.87 | 2.15 | 3.29 | .79 | 1.36 | 1.57 |
| 12 | 0–2 | −.30 | 3.89 | 5.93 | −1.99 | 4.34 | 7.55 | .92 | 1.20 | 1.79 |
| 13 | 0–3 | .31 | 5.26 | 7.57 | −3.72 | 6.51 | 11.43 | .92 | 1.11 | 1.37 |
| 14 | 0–4 | 1.36 | 6.59 | 9.02 | −5.43 | 8.43 | 14.32 | .94 | 1.16 | 1.29 |
| 15 | 0–5 | 2.32 | 8.19 | 10.56 | −7.51 | 10.55 | 17.46 | .93 | 1.14 | 1.26 |
| | | National Defense Expenditures (DEF) | | | | | | | | |
| 16 | 0–1 | −.09 | 2.18 | 2.54 | −.07 | 1.44 | 2.00 | .85 | .98 | 1.20 |
| 17 | 0–2 | −.28 | 2.89 | 3.65 | .56 | 2.28 | 3.08 | .84 | 1.04 | 1.19 |
| 18 | 0–3 | −.49 | 3.75 | 4.52 | −1.49 | 3.13 | 4.09 | .74 | .92 | 1.15 |
| 19 | 0–4 | −.65 | 4.03 | 4.76 | −2.14 | 4.34 | 5.23 | .79 | 1.01 | 1.25 |
| 20 | 0–5 | −.95 | 5.84 | 6.83 | −3.64 | 5.45 | 7.07 | .67 | .80 | .99 |

*Note:* See table 1.21.

### 1.5.2 Sims's Probabilistic Forecasts

In addition to outside econometric model forecasts, we wished to compare the results of the NBER-ASA surveys to outside time-series forecasts. We are indebted to Chris Sims for data on predictions from both sophisticated BVAR and univariate ARIMA models.

Recall that our own BVAR model used earlier in this paper includes RGNP, IPD, TBER, M2, and LI plus the variable predicted (if not one of the above). The Sims model includes the first three variables in our set plus six others:

**Table 1.23** **Michigan (RSQE) Econometric Forecasts and NBER-ASA Forecasts of the Unemployment Rate, Treasury-Bill Rate, and Corporate Bond Yield, by Target Quarter, 1968–90 and 1981–90**

| | | Michigan Forecasts | | | Group Mean Forecasts | | | RMSE Ratios i/Mi | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Row | Target Quarter | M (1) | MAE (2) | RMSE (3) | M (4) | MAE (5) | RMSE (6) | $Q_1$ (7) | MD (8) | $Q_3$ (9) |
| | | | | Unemployment Rate, 1968–90 (UR) | | | | | | |
| 1 | 1 | .05 | .14 | .17 | .02 | .13 | .17 | 1.07 | 1.32 | 1.72 |
| 2 | 2 | .08 | .33 | .44 | .05 | .33 | .43 | 1.03 | 1.15 | 1.35 |
| 3 | 3 | .05 | .49 | .67 | .03 | .51 | .68 | 1.01 | 1.16 | 1.32 |
| 4 | 4 | −.01 | .58 | .78 | .00 | .61 | .85 | 1.08 | 1.21 | 1.38 |
| 5 | 5 | −.11 | .69 | .93 | −.02 | .71 | .96 | 1.01 | 1.18 | 1.33 |
| | | | | Treasury-Bill Rate, 1981–90 (TBR) | | | | | | |
| 6 | 1 | −.04 | .24 | .31 | .01 | .15 | .20 | .96 | 1.17 | 1.58 |
| 7 | 2 | −.05 | .79 | 1.07 | .15 | .69 | .91 | .85 | .96 | 1.19 |
| 8 | 3 | −.01 | 1.13 | 1.39 | .38 | 1.11 | 1.40 | .97 | 1.11 | 1.32 |
| 9 | 4 | .07 | 1.37 | 1.64 | .62 | 1.45 | 1.80 | 1.05 | 1.25 | 1.39 |
| 10 | 5 | .21 | 1.67 | 1.90 | .87 | 1.72 | 2.16 | 1.06 | 1.25 | 1.52 |
| | | | | Corporate Bond Yield, 1981–90 (CBY) | | | | | | |
| 11 | 1 | −.44 | .48 | .63 | −.19 | .31 | .38 | .73 | .97 | 1.26 |
| 12 | 2 | −.31 | .64 | .81 | −.07 | .66 | .83 | 1.17 | 1.28 | 1.55 |
| 13 | 3 | −.20 | .84 | 1.08 | .16 | 1.05 | 1.25 | 1.23 | 1.36 | 1.50 |
| 14 | 4 | −.12 | 1.17 | 1.43 | .37 | 1.32 | 1.53 | 1.08 | 1.23 | 1.39 |
| 15 | 5 | −.03 | 1.39 | 1.68 | .57 | 1.48 | 1.74 | 1.02 | 1.12 | 1.32 |

*Note:* See table 1.21.

M1, UR, NFI, the S&P 500 stock price index, a commodity price index, and the trade-weighted value of the dollar.[13] It is a nine-variable, five-lag model, whereas ours is a five- or six-variable, six-lag model.

Sims's model is an extension of the model constructed in 1980 and used in quarterly forecasting during 1980–86 by Litterman (1986). It is three variables larger than the original Litterman model, and it allows time variation in coefficients, predictable time variation in forecast error variance, and nonnormality in disturbances (Sims 1989). The modifications give rise to nonnormal, nonlinear models and hence to considerable complications in estimation and analysis (Sims and Todd 1991). The Sims model (like our own BVAR) forecasts are simulations of real-time forecasts in that they use only data from time periods before the periods to be predicted. But, for several reasons, including

13. The data are generally expressed in log-level form, except for TBR, which was not logged.

the use of current versions of the data, they are far from being true ex ante forecasts (again, the same applies to our BVAR as well).

In evaluating the BVAR forecasts (both Sims's and our own), we used the current data, which is consistent with their construction and believed to be fair. Use of preliminary figures would have resulted in finding larger errors.

Again, like for our own BVAR (see table 1.13 and text above), the comparisons of the Sims model forecasts with the NBER-ASA survey forecasts for GNP, RGNP, and IPD are presented in two variants, A and B (table 1.24). For reasons already explained, variant A favors the real-time predictions that incorporate contemporary news evaluations, while variant B favors the predictions based on the ex post constructed time-series models.

Using variant A, Sims's forecasts (S) are found to have on the whole larger errors than the group mean forecasts from the NBER-ASA surveys for both GNP and RGNP (table 1.24, rows 1–5 and 11–15; cf. cols. 1–3 and 4–6). The corresponding ratios $RMSE_i/RMSE_s$ are relatively low, approaching 1.00 only for $Q_3$ (cols. 7–9), which means that most individual forecasts from the surveys are more accurate than the Sims model forecasts. In contrast, the Sims forecasts are considerably more accurate than the group mean forecasts for IPD inflation, and here the RMSE ratios i/S mostly exceed 1.00, even for $Q_1$ (rows 21–25).

Using variant B as a criterion (rows 6–10, 16–20, and 26–30), we still see the group mean forecasts as retaining on balance an advantage over the Sims forecasts for GNP, but it is a much-reduced advantage and one essentially limited to the longer spans. For RGNP, the NBER-ASA consensus predictions are somewhat more accurate than the Sims model predictions for the spans 0–4 and 0–5, whereas the opposite is true for the shorter spans. For IPD, the Sims forecasts have smaller errors throughout. (Compare cols. 1–3 for variant B with the corresponding entries in cols. 4–6.) Looking at the RMSE ratios i/S (cols. 7–9), we find them to exceed 1.00, that is, to favor the Sims model, for GNP at $Q_3$ only, for RGNP at MD and $Q_3$, and for IPD at $Q_1$, MD, and $Q_3$.

Interestingly, the original Litterman BVAR performed relatively well for real GNP and unemployment but worse for IPD, which motivated both Litterman and Sims to make changes designed to improve their inflation forecasts. But simulations disclosed "a tendency for improvements in the retrospective forecast performance of the BVAR model for inflation to be accompanied by deterioration in its performance for real variables" (Sims 1989, 1). A similar trade-off was observed in working with our own BVAR.

According to the measures in table 1.25 (based on variant A only), most of the NBER-ASA survey forecasts for the unemployment rate (1968–90), the Treasury-bill rate (1981–90), and the rate of growth in real nonresidential fixed investment (1981–90) exceeded the corresponding Sims model forecasts considerably in overall accuracy. This can be concluded from both the com-

**Table 1.24 Sims Model Forecasts (Two Variants) and NBER-ASA Survey Forecasts of Nominal and Real GNP Growth and IPD Inflation, by Span, 1968–90 and 1981–90**

| | | Sims Model Forecasts (S) | | | Group Mean Forecasts | | | RMSE Ratios i/S | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Row | Span (Qs) | M (1) | MAE (2) | RMSE (3) | M (4) | MAE (5) | RMSE (6) | $Q_1$ (7) | MD (8) | $Q_3$ (9) |
| | | Gross National Product, 1968–90 (GNP): Variant A | | | | | | | | |
| 1 | 0–1 | .01 | .86 | 1.09 | −.08 | .49 | .64 | .51 | .66 | .89 |
| 2 | 0–2 | .01 | 1.31 | 1.68 | −.10 | .84 | 1.11 | .62 | .79 | .94 |
| 3 | 0–3 | −.05 | 1.87 | 2.34 | −.07 | 1.22 | 1.61 | .62 | .79 | .93 |
| 4 | 0–4 | −.11 | 2.32 | 2.93 | −.02 | 1.56 | 2.06 | .65 | .79 | .98 |
| 5 | 0–5 | −.20 | 2.74 | 3.48 | −.09 | 1.91 | 2.51 | .65 | .80 | 1.08 |
| | | GNP, 1968–90: Variant B | | | | | | | | |
| 6 | 0–1 | 0 | 0 | 0 | | | | N.A. | N.A. | N.A. |
| 7 | 0–2 | .00 | .85 | 1.08 | | | | .70 | .84 | .98 |
| 8 | 0–3 | .02 | 1.29 | 1.66 | | | | .81 | .93 | 1.10 |
| 9 | 0–4 | .08 | 1.85 | 2.33 | | | | .74 | .88 | 1.09 |
| 10 | 0–5 | .15 | 2.29 | 2.90 | | | | .78 | .88 | 1.10 |
| | | GNP in Constant Dollars, 1968–90 (RGNP): Variant A | | | | | | | | |
| 11 | 0–1 | .03 | .78 | .99 | −.02 | .50 | .64 | .63 | .77 | .98 |
| 12 | 0–2 | .05 | 1.18 | 1.50 | .02 | .83 | 1.11 | .73 | .86 | 1.06 |
| 13 | 0–3 | .04 | 1.68 | 2.12 | .16 | 1.17 | 1.61 | .75 | .89 | 1.05 |
| 14 | 0–4 | .02 | 2.10 | 2.66 | .33 | 1.42 | 2.05 | .79 | .93 | 1.08 |
| 15 | 0–5 | −.03 | 2.54 | 3.15 | .40 | 1.70 | 2.47 | .80 | .96 | 1.12 |
| | | RGNP, 1968–90: Variant B | | | | | | | | |
| 16 | 0–1 | 0 | 0 | 0 | | | | N.A. | N.A. | N.A. |
| 17 | 0–2 | −.02 | .79 | 1.00 | | | | .81 | .96 | 1.11 |
| 18 | 0–3 | −.03 | 1.19 | 1.51 | | | | .99 | 1.12 | 1.27 |
| 19 | 0–4 | −.03 | 1.70 | 2.14 | | | | .97 | 1.12 | 1.30 |
| 20 | 0–5 | .00 | 2.11 | 2.68 | | | | 1.01 | 1.13 | 1.36 |
| | | Implicit Price Deflator, 1968–90 (IPD): Variant A | | | | | | | | |
| 21 | 0–1 | .01 | .30 | .38 | −.04 | .27 | .34 | .71 | .88 | 1.17 |
| 22 | 0–2 | .04 | .54 | .68 | −.12 | .55 | .70 | 1.81 | .97 | 1.17 |
| 23 | 0–4 | .08 | .75 | .95 | −.22 | .84 | 1.13 | 1.24 | 1.46 | 1.73 |
| 24 | 0–4 | .12 | 1.01 | 1.25 | −.34 | 1.21 | 1.64 | 1.35 | 1.58 | 1.85 |
| 25 | 0–5 | .16 | 1.29 | 1.59 | −.37 | 1.63 | 2.23 | 1.40 | 1.68 | 1.99 |
| | | IPD, 1968–90: Variant B | | | | | | | | |
| 26 | 0–1 | 0 | 0 | 0 | | | | N.A. | N.A. | N.A. |
| 27 | 0–2 | −.02 | .31 | .39 | | | | 1.31 | 1.52 | 1.81 |
| 28 | 0–3 | −.05 | .55 | .69 | | | | 1.43 | 1.63 | 1.91 |

*(continued)*

**Table 1.24** (continued)

| | | Sims Model Forecasts (S) | | | Group Mean Forecasts | | | RMSE Ratios i/S | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Row | Span (Qs) | M (1) | MAE (2) | RMSE (3) | M (4) | MAE (5) | RMSE (6) | $Q_1$ (7) | MD (8) | $Q_3$ (9) |
| 29 | 0–4 | −.09 | .76 | .96 | | | | 1.51 | 1.68 | 1.90 |
| 30 | 0–5 | −.14 | 1.02 | 1.26 | | | | 1.47 | 1.71 | 1.98 |

*Note:* Sims's model is a nine-variable, five-lag quarterly probabilistic model (see the text for more detail). The Sims forecasts contain no gaps and refer to the same periods as those covered by the NBER-ASA survey forecasts (individual and group means). The entries in cols. 7–9 represent ratios $RMSE_i/RMSE_S$, where the subscript i refers to individual forecasts from the surveys and the subscript S refers to the Sims model forecasts. $Q_1$ and $Q_3$ denote the lower- and upper-quartile ratios, and MD denotes the median ratio. Variant A assumes that the last known values of the variables to the predicted refer to the quarter $t - 1$ (denoted 0); variant B assumes that they refer to the current quarter $t$ (denoted 1). N.A. = not available.

parisons with group mean predictions from the surveys (cf. cols. 1–3 and 4–6) and the low i/S ratios (cols. 7–9).

The Sims model and our own BVAR forecasts have errors of generally similar orders of magnitude. The Sims predictions are more accurate for GNP and IPD, less accurate for RGNP and UR. The results for NFI and TBR are mixed (favoring Sims at the two longest horizons only).[14]

### 1.5.3 Univariate Time-Series Models

Predictions from ARIMA models make popular benchmarks for evaluating forecasters' performance. We use ARIMA as specified in Sims and Todd (1991), where they are reported to have worked well relative to the Simsian BVAR for financial variables and business fixed investment in 1980–90 (pp. 9–10). However, our measures show that, throughout, Sims's BVAR forecasts had smaller overall errors than the corresponding ARIMA forecasts, whether the comparisons cover the variants A or the variants B (cf. tables 1.24 and 1.25, cols. 1–3, with tables 1.26 and 1.27, cols. 2–4).

The results of comparing the NBER-ASA survey forecasts with their counterparts of the Sims-Todd ARIMA type are less clear-cut. Most of the forecasters did better than the time-series models according to the variant A calculations, as is evident from the individual-to-ARIMA (i/Ar) ratios in columns 5–7 of tables 1.26 and 1.27. But, when variant B is used, the forecasters are no longer clearly ahead for RGNP and fall somewhat behind for IPD (table 1.26, rows 16–20 and 26–30).

Beginning in 1976:II, Charles Nelson has produced ARIMA forecasts of rates of change in nominal and real GNP and the implicit price deflator synchronously with other real-time forecasts, updating them each quarter on the

14. For the RMSEs of the BVAR forecasts, see table 1.13, col. 3 (GNP, RGNP, IPD), and tables 1.19, 1A.6, and 1A.8, col. 5 (UR, NFI, and TBR, respectively).

**Table 1.25** **Sims Model Forecasts (Variant A) and NBER-ASA Survey Forecasts of the Unemployment Rate, the Treasury-Bill Rate, and Growth in Real Nonresidential Investment, by Target Quarter or Span, 1968–90 and 1981–90**

| Row | Target Quarter or Span (Qs) | Sims Model Forecasts | | | Group Mean Forecasts | | | RMSE Ratios i/S | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | M (1) | MAE (2) | RMSE (3) | M (4) | MAE (5) | RMSE (6) | $Q_1$ (7) | MD (8) | $Q_3$ (9) |
| | | Unemployment Rate, 1968–90 (UR) | | | | | | | | |
| 1 | 1 | .09 | .39 | .55 | .03 | .13 | .16 | .35 | .45 | .55 |
| 2 | 2 | .14 | .56 | .79 | .04 | .32 | .41 | .54 | .65 | .81 |
| 3 | 3 | .18 | .76 | 1.03 | −.00 | .49 | .65 | .59 | .71 | .93 |
| 4 | 4 | .21 | .95 | 1.23 | −.08 | .63 | .86 | .64 | .78 | .94 |
| 5 | 5 | .23 | 1.10 | 1.40 | −.10 | .73 | 1.00 | .66 | .79 | .98 |
| | | Treasury-Bill Rate, 1981–90 (TBR) | | | | | | | | |
| 6 | 1 | −.34 | 1.27 | 1.57 | .01 | .15 | .20 | .20 | .24 | .29 |
| 7 | 2 | −.54 | 1.47 | 1.84 | .13 | .68 | .90 | .52 | .62 | .70 |
| 8 | 3 | −.60 | 1.69 | 2.13 | .35 | 1.09 | 1.38 | .62 | .71 | .82 |
| 9 | 4 | −.71 | 1.96 | 2.48 | .61 | 1.41 | 1.77 | .67 | .76 | .86 |
| 10 | 5 | −.86 | 2.19 | 2.69 | 1.07 | 1.87 | 2.49 | .75 | .86 | .97 |
| | | Nonresidential Fixed Investment, 1981–90 (NFI) | | | | | | | | |
| 11 | 0–1 | −.16 | 2.31 | 2.93 | −.45 | 1.61 | 2.01 | .72 | .85 | 1.01 |
| 12 | 0–2 | −.31 | 3.61 | 4.16 | −.88 | 2.67 | 3.43 | .63 | .72 | 1.01 |
| 13 | 0–3 | −.74 | 5.23 | 6.05 | −1.19 | 3.93 | 4.99 | .60 | .67 | .85 |
| 14 | 0–4 | −.90 | 4.50 | 6.69 | −1.74 | 5.51 | 6.89 | .60 | .65 | .93 |
| 15 | 0–5 | −1.63 | 4.95 | 7.31 | −2.31 | 7.29 | 8.69 | .54 | .65 | .75 |

*Note:* See table 1.24.

**Table 1.26 ARIMA Model Forecasts (Two Variants) and NBER-ASA Survey Forecasts of Nominal and Real GNP Growth and IPD Inflation, by Span, 1968–90**

| Row | Span (Qs) | ARIMA Model (Ar) (1) | ARIMA Forecasts M (2) | ARIMA Forecasts MAE (3) | ARIMA Forecasts RMSE (4) | RMSE Ratios i/Ar $Q_1$ (1) | RMSE Ratios i/Ar MD (6) | RMSE Ratios i/Ar $Q_3$ (7) |
|---|---|---|---|---|---|---|---|---|
| | | Gross National Product (GNP): Variant A | | | | | | |
| 1 | 0–1 | N.A. | −.11 | .95 | 1.18 | .48 | .61 | .82 |
| 2 | 0–2 | | −.29 | 1.64 | 2.05 | .53 | .64 | .74 |
| 3 | 0–3 | | −.55 | 2.51 | 3.04 | .51 | .61 | .73 |
| 4 | 0–4 | | −.85 | 3.32 | 4.00 | .51 | .58 | .69 |
| 5 | 0–5 | | −1.19 | 4.11 | 4.96 | .49 | .61 | .74 |
| | | GNP: Variant B | | | | | | |
| 6 | 0–1 | N.A. | 0 | 0 | 0 | N.A. | N.A. | N.A. |
| 7 | 0–2 | | .12 | .94 | 1.18 | .64 | .77 | .86 |
| 8 | 0–3 | | .32 | 1.62 | 2.02 | .71 | .81 | .95 |
| 9 | 0–4 | | .59 | 2.48 | 3.01 | .61 | .71 | .87 |
| 10 | 0–5 | | .91 | 3.28 | 3.96 | .59 | .74 | .94 |
| | | Gross National Product in Constant Dollars (RGNP): Variant A | | | | | | |
| 11 | 0–1 | 1,1,0 | −.06 | .80 | 1.03 | .60 | .74 | .94 |
| 12 | 0–2 | | −.14 | 1.33 | 1.68 | .67 | .79 | .94 |
| 13 | 0–3 | | −.27 | 1.91 | 2.33 | .71 | .82 | 1.00 |
| 14 | 0–4 | | −.40 | 2.37 | 2.88 | .77 | .86 | 1.05 |
| 15 | 0–5 | | −.54 | 2.82 | 3.39 | .76 | .93 | 1.11 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | RGNP: Variant B | | | | | | |
| 16 | 0–1 | 1,1,0 | 0 | 0 | 0 | N.A. | N.A. | N.A. |
| 17 | 0–2 | | .06 | .81 | 1.04 | .78 | .93 | 1.08 |
| 18 | 0–3 | | .16 | 1.33 | 1.68 | .91 | 1.04 | 1.21 |
| 19 | 0–4 | | .28 | 1.93 | 2.35 | .93 | 1.07 | 1.29 |
| 20 | 0–5 | | .43 | 2.39 | 2.91 | .99 | 1.11 | 1.34 |
| | | Implicit Price Deflator (IPD): Variant A | | | | | | |
| 21 | 0–1 | 1,1,2 | .05 | .38 | .50 | .72 | .93 | 1.11 |
| 22 | 0–2 | | .15 | .80 | 1.00 | .71 | .87 | 1.10 |
| 23 | 0–4 | | .29 | 1.27 | 1.60 | .68 | .90 | 1.11 |
| 24 | 0–4 | | .49 | 1.84 | 2.29 | .65 | .93 | 1.16 |
| 25 | 0–5 | | .74 | 2.47 | 3.07 | .63 | 1.02 | 1.28 |
| | | IPD: Variant B | | | | | | |
| 26 | 0–1 | 1,1,2 | 0 | 0 | 0 | N.A. | N.A. | N.A. |
| 27 | 0–2 | | −.05 | .39 | .51 | .94 | 1.10 | 1.39 |
| 28 | 0–3 | | −.16 | .80 | 1.01 | .97 | 1.17 | 1.37 |
| 29 | 0–4 | | −.32 | 1.27 | 1.62 | .88 | 1.11 | 1.33 |
| 30 | 0–5 | | −.53 | 1.86 | 2.32 | .79 | 1.14 | 1.40 |

*Note:* N.A. = not available (forecasts obtained from those for RGNP and IPD). The specifications of the ARIMA models are as in Sims and Todd (1991, table 1). For more detail, see Sims and Todd (1991, 3–4). The entries in cols. 5–6 represent ratios $RMSE_i/RMSE_{Ar}$, where the subscript i refers to individual forecasts from the NBER-ASA surveys and the subscript Ar refers to the ARIMA model forecasts.

Table 1.27 **ARIMA Model Forecasts (Variant A) and NBER-ASA Survey Forecasts of the Unemployment Rate, the Treasury-Bill Rate, and Growth in Real Nonresidential Investment, by Target Quarter or Span, 1968–90 and 1981–90**

| Row | Target Quarter or Span (Qs) | ARIMA Model (Ar) (1) | ARIMA Forecasts M (2) | ARIMA Forecasts MAE (3) | ARIMA Forecasts RMSE (4) | RMSE Ratios i/Ar $Q_1$ (5) | RMSE Ratios i/Ar MD (6) | RMSE Ratios i/Ar $Q_3$ (7) |
|---|---|---|---|---|---|---|---|---|
| | | | Unemployment Rate (UR) | | | | | |
| 1 | 1 | 1,1,0 | −.25 | .45 | .65 | .30 | .36 | .44 |
| 2 | 2 | | −.44 | .67 | .96 | .44 | .52 | .60 |
| 3 | 3 | | −.63 | .87 | 1.26 | .50 | .58 | .68 |
| 4 | 4 | | −.80 | 1.07 | 1.50 | .56 | .63 | .74 |
| 5 | 5 | | −.94 | 1.21 | 1.67 | .58 | .66 | .76 |
| | | | Treasury-Bill Rate (TBR) | | | | | |
| 6 | 1 | 0,1,1 | −.39 | 1.37 | 1.96 | .15 | .19 | .26 |
| 7 | 2 | | −.66 | 1.60 | 2.14 | .42 | .49 | .61 |
| 8 | 3 | | −.80 | 1.88 | 2.59 | .53 | .60 | .72 |
| 9 | 4 | | −.96 | 2.26 | 3.19 | .57 | .65 | .72 |
| 10 | 5 | | −1.19 | 2.51 | 3.49 | .59 | .69 | .78 |
| | | | Nonresidential Fixed Investment (NFI) | | | | | |
| 11 | 0–1 | 1,1,0 | −.07 | 1.63 | 2.37 | .58 | .84 | 1.21 |
| 12 | 0–2 | | −.16 | 3.32 | 4.27 | .65 | .78 | .98 |
| 13 | 0–3 | | −.22 | 5.10 | 6.26 | .58 | .64 | .81 |
| 15 | 0–4 | | −.22 | 6.76 | 8.11 | .50 | .59 | .65 |
| 16 | 0–5 | | −.16 | 8.19 | 9.69 | .40 | .48 | .56 |

*Note:* See table 1.26.

**Table 1.28** **Joutz Model Forecasts and NBER-ASA Survey Forecasts of Nominal and Real GNP Growth and IPD Inflation, by Span, 1976–90**

| | | Joutz Model Forecasts (J) | | | Group Mean Forecasts (g) | | | RMSE Ratios i/J | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Row | Span (Qs) | M (1) | MAE (2) | RMSE (3) | M (4) | MAE (5) | RMSE (6) | $Q_1$ (7) | MD (8) | $Q_3$ (9) |
| | | Gross National Product (GNP) | | | | | | | | |
| 1 | 0–1 | .04 | .72 | .96 | −.05 | .55 | .70 | .73 | .89 | 1.25 |
| 2 | 0–2 | .09 | 1.21 | 1.52 | −.02 | .89 | 1.16 | .85 | 1.08 | 1.26 |
| 3 | 0–3 | .16 | 1.65 | 2.08 | .08 | 1.30 | 1.69 | 1.00 | 1.18 | 1.64 |
| 4 | 0–4 | .28 | 2.05 | 2.58 | .18 | 1.63 | 2.15 | .94 | 1.14 | 1.39 |
| | | Gross National Product in Constant Dollars (RGNP) | | | | | | | | |
| 5 | 0–1 | .02 | .64 | .85 | −.08 | .53 | .65 | .86 | 1.07 | 1.40 |
| 6 | 0–2 | .08 | 1.05 | 1.31 | −.10 | .81 | 1.03 | .84 | 1.03 | 1.45 |
| 7 | 0–3 | .18 | 1.38 | 1.68 | −.05 | 1.03 | 1.35 | .95 | 1.21 | 1.55 |
| 8 | 0–4 | .30 | 1.53 | 1.90 | −.04 | 1.12 | 1.57 | .89 | 1.12 | 1.51 |
| | | Implicit Price Deflator (IPD) | | | | | | | | |
| 9 | 0–1 | .01 | .29 | .37 | .04 | .24 | .30 | .86 | 1.07 | 1.40 |
| 10 | 0–2 | .00 | .53 | .65 | .09 | .43 | .52 | .78 | .96 | 1.30 |
| 11 | 0–3 | −.03 | .73 | .96 | .15 | .61 | .75 | .80 | .98 | 1.21 |
| 12 | 0–4 | −.04 | 1.05 | 1.33 | .21 | .89 | 1.06 | .79 | .97 | 1.28 |

announcement of the first preliminary numbers for the preceding quarter. Comparisons with five econometric models for the period 1976:II–1982:IV have shown these ex ante "benchmark" forecasts to be of competitive accuracy (Nelson 1984). Since 1988, Frederick Joutz has been preparing the ARIMA forecasts on a current basis (the same way as Nelson had before), and he kindly let us have the results for the purposes of a comparative analysis.

Table 1.28 shows that the NBER-ASA group mean forecasts (g) were on average consistently more accurate than the Joutz ARIMA (J) forecasts (cf. cols. 2–5 and 3–6). The RMSE ratios g/J rose with the span from 0.73 to 0.88 for GNP and from 0.76 to 0.83 for RGNP; they varied irregularly between 0.78 and 0.81 for IPD. The RMSE ratios i/J (cols. 7–9) average 0.8–0.9 for $Q_1$, 1.0–1.1 for MD, and 1.3–1.5 for $Q_3$. Our analysis confirms the findings that these ARIMA forecasts are indeed competitive and that their relative accuracy tends to improve with their horizon for GNP and RGNP (but not for IPD, where they are weakest).

## 1.6 A General Evaluation and Conclusions

In presenting and discussing more than thirty tables on multiperiod quarterly forecasts for a score of variables by a total of more than one hundred individuals, we had to make some hard choices about which problems to confront and which measures to use. Forecasts for two-thirds of the time series covered were treated less comprehensively and relegated to an appendix, to make the paper easier to read. Even so, the inevitable abundance of detail risks obscuring the overall picture. Therefore, lest we miss the forest for the trees, a statement of general findings, conclusions, and qualifications is very necessary at this point.

1. The distributions of the error statistics show that there is much dispersion across the forecasts, which typically increases with the length of the predictive horizon. Forecasters differ in many respects, and so do their products. The idea that a close "consensus" persists, that is, that current matched forecasts are generally all alike, is a popular fiction. The differentiation of the forecasts usually involves much more than the existence of just a few outliers. However, it is also true that forecasters depend on common information, interact, and influence each other. This naturally induces some common trends. The more independent information the individuals possess, the more their predictions can differ. Thus, a clustering of forecasts could be due either to genuine agreement or to common ignorance, while dissent may reflect uncertainty.[15]

2. Errors of the average change forecasts cumulate over the spans 0–1, . . . , 0–5 with great regularity for a variety of time series. To a large extent,

15. Compare Zarnowitz and Lambros (1987), a study that compares the point and probabilistic forecasts from the NBER-ASA surveys. Time and space restrictions prevented us from including in this paper the survey responses to questions on the probabilities of alternative GNP and IPD outcomes and turning points. See also Braun and Yaniv (1991).

this occurs because of the progression to larger changes in the corresponding actual values. But the errors of marginal change and level forecasts, too, often increase with the distance to the target quarter, although by much smaller margins and with much less regularity. As might be expected, the further out in the future the target, the less can be inferred about it from the past, and the worse it is usually forecast. The less random and more predictable the series, the better this rule holds, in the sense that the forecasts will be more forward looking and more appropriately differentiated with the distance to the target period.[16]

3. Macroeconomic variables differ greatly in the extent to which they can be forecast. The more persistent (autocorrelated) series are, of course, more accurately predicted than series with high random variability. Thus, real GNP and consumption are far easier to forecast than residential investment and, especially, change in business inventories. Inflation was underestimated and poorly predicted by most forecasters most of the time. Negative correlations between RGNP and IPD forecast errors have long been observed (see Zarnowitz 1979, table 4 and text), and offsetting performance for inflation and real variables appears to be frequently encountered in studies of forecasting methods and results.

4. A comparison of the summary measures of error for 1968:IV–1979:III and 1979:IV–1990:I reveals no large and systematic differences that would indicate either deterioration or improvement in the overall performance of the respondents to the NBER-ASA surveys. The accuracy of GNP forecasts may have decreased somewhat, but that of inflation forecasts increased. The 1970s and the 1980s differed significantly in a number of economically important dimensions, but it is difficult to say that either subperiod presented the forecasts with definitely greater problems than the other. Each experienced two business recessions, which is noted because previous research has shown that turning-point errors played a major role in downgrading the forecasting records (for a recent summary, see Zarnowitz 1992).

5. Group mean forecasts are generally much more accurate than the majority of individual forecasts. These consensus predictions are computed by simple averaging across the corresponding responses to each successive survey; we made no effort to use other than equal weighting. This paper, then, provides many examples of the rule that combining forecasts often results in substantial improvements. The method is very accessible and inexpensive. The gains are enhanced by the diversification of the forecasts that are combined; for example, our group mean forecasts should be better the more different and complementary the information embodied in their components. For some variables and periods, the combinations work much better than for oth-

16. It should be noted that annual forecasts are generally more accurate than all but the very short quarterly forecasts, owing to cancellation of errors for the quarters within the year (Zarnowitz 1979). In this paper, annual forecasts are not considered.

**Table 1.29** **Nine Sets of Forecasts Ranked According to Their Average RMSEs, Three Variables, 1968:IV–1990:I**

| | | Gross National Product (GNP) | | GNP in Constant Dollars (RGNP) | | Implicit Price Deflator (IPD) | |
|---|---|---|---|---|---|---|---|
| Row | 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 variant A | 2.69 | 8 | 1.90 | 6 | 1.62 | 8 |
| 5 | BVAR variant B | 1.89 | 3 | 1.40 | 1 | 1.03 | 3 |
| 6 | Sims variant A | 2.30 | 7 | 2.08 | 8 | .97 | 2 |
| 7 | Sims variant B | 1.594 | 2 | 1.47 | 2 | .66 | 1 |
| 8 | Sims-Todd ARIMA variant A | 3.05 | 9 | 2.26 | 9 | 1.69 | 9 |
| 9 | Sims-Todd ARIMA variant B | 2.03 | 6 | 1.60 | 4 | 1.09 | 4 |

*Source:* Row 1 is based on entries in table 1.10, row 3, cols. 1–5; row 2 on table 1.9, col. 6; row 3 on table 1.21, col. 3; rows 4 and 5 on table 1.13, col. 3; rows 6 and 7 on table 1.24, col. 3; and rows 8 and 9 on table 1.26, col. 4.

*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, the largest ARMSE 9, for each of the three variables.

ers. In principle, one would prefer to combine the information in a single model rather than combining the forecasts. In practice, the latter will typically be much easier.

6. Consider first comparisons with time-series models constructed on the assumption that the last-known values of the variables concerned refer to the prior quarter $t - 1$ (variant A). The assumption is certainly valid for the quarterly variables in the real-time forecasts, but it results in some bias against the time-series forecasts. Table 1.29 sums up the evidence in the form of the RMSEs averaged across spans. For the subset consisting of the median individual and the consensus forecasts from the NBER-ASA surveys, the Michigan econometric model, our BVAR variant A model, the Sims variant A probabilistic model, and the Sims-Todd ARIMA variant A model (rows 1–4, 6, and 8), the consensus (group mean) survey forecasts rank first for GNP and RGNP and second for IPD (following the Sims [variant A] model).

7. The alternative assumption, that the last-known values of the variables refer to the current quarter $t$ (variant B), is rather strongly biased in favor of the ex post forecasts with time-series models. The average root mean square errors (ARMSEs) are all much lower for the variant B predictions than for their variant A counterparts (cf. rows 4, 6, and 8 with rows 5, 7, and 9). When all nine sets of forecasts listed in table 1.29 are considered, the Sims variant B model ranks second, second, and first for GNP, RGNP, and IPD, respec-

**Table 1.30** **Six Sets of Forecasts Ranked According to Their RMSEs Averaged across Spans, Twenty-one Variables, 1968:IV–1990:I, 1968:IV–1981:II, and 1981:II–1990:I**

| | | Average Root Mean Square Error (ARMSE) and the Corresponding Rank[b] | | | | | |
|---|---|---|---|---|---|---|---|
| Row | Variable[a] | NBER-ASA Surveys Median Individual Forecast (1) | Group (Consensus) Forecast (2) | BVAR Model Forecast (3) | Michigan (RSQE) Forecast (4) | Sims Probabilistic Model Forecast (5) | ARIMA Model Forecast (6) |
| | | 1968:IV–1990:I | | | | | |
| 1 | GNP | 1.90 | 1.59 | 2.69 | 1.98 | 2.30 | 3.05 |
| | | (2) | (1) | (5) | (3) | (4) | (6) |
| 2 | RGNP | 1.94 | 1.58 | 1.90 | 1.87 | 2.08 | 2.26 |
| | | (4) | (1) | (3) | (2) | (5) | (6) |
| 3 | IPD | 1.53 | 1.21 | 1.62 | 1.42 | .97 | 1.69 |
| | | (4) | (2) | (5) | (3) | (1) | (6) |
| 4 | IP | 4.02 | 3.56 | 3.42 | N.A. | N.A. | N.A. |
| | | (3) | (2) | (1) | | | |
| 5 | CP | 18.38 | 17.13 | 13.71 | N.A. | N.A. | N.A. |
| | | (3) | (2) | (1) | | | |
| 6 | UR | .71 | .62 | .61 | .60 | 1.00 | 1.21 |
| | | (4) | (3) | (2) | (1) | (5) | (6) |
| 7 | HS | .33 | .29 | .26 | N.A. | N.A. | N.A. |
| | | (3) | (2) | (1) | | | |
| | | 1968:IV–1981:II | | | | | |
| 8 | CD | 5.65 | 5.05 | 5.93 | N.A. | N.A. | N.A. |
| | | (2) | (1) | (3) | | | |
| 9 | PE | 11.49 | 11.08 | 4.35 | N.A. | N.A. | N.A. |
| | | (3) | (2) | (1) | | | |
| 10 | DEF | 3.92 | 3.34 | 8.19 | 4.46 | N.A. | N.A. |
| | | (2) | (1) | (4) | (3) | | |
| 11 | CBI | 11.38 | 12.38 | 13.94 | N.A. | N.A. | N.A. |
| | | (1) | (2) | (3) | | | |
| | | 1981:III–1990:I | | | | | |
| 12 | RCBI | 19.69 | 18.96 | 26.69 | N.A. | N.A. | N.A. |
| | | (2) | (1) | (3) | | | |
| 13 | NX | 47.29 | 44.19 | 21.02 | N.A. | N.A. | N.A. |
| | | (3) | (2) | (1) | | | |
| 14 | PCE | 1.30 | 1.24 | 5.69 | 1.13 | N.A. | N.A. |
| | | (3) | (2) | (4) | (1) | | |
| 15 | NFI | 6.06 | 1.11 | 5.25 | 5.80 | 5.43 | 6.14 |
| | | (5) | (1) | (2) | (4) | (3) | (6) |
| 16 | RFI | 9.59 | 1.22 | 8.70 | 10.81 | N.A. | N.A. |
| | | (3) | (1) | (2) | (4) | | |
| 17 | FGP | 4.96 | 1.38 | 8.54 | N.A. | N.A. | N.A. |
| | | (2) | (1) | (3) | | | |

*(continued)*

Table 1.30 (continued)

| Row | Variable[a] | NBER-ASA Surveys Median Individual Forecast (1) | Group (Consensus) Forecast (2) | BVAR Model Forecast (3) | Michigan (RSQE) Forecast (4) | Sims Probabilistic Model Forecast (5) | ARIMA Model Forecast (6) |
|---|---|---|---|---|---|---|---|
| | | Average Root Mean Square Error (ARMSE) and the Corresponding Rank[b] | | | | | |
| 18 | SLGP | 1.47 | .94 | 1.27 | N.A. | N.A. | N.A. |
| | | (3) | (1) | (2) | | | |
| 19 | CPI | 1.19 | .52 | .76 | N.A. | N.A. | N.A. |
| | | (3) | (1) | (2) | | | |
| 20 | TBR | 1.71 | 1.35 | 2.03 | 1.26 | 2.14 | 2.67 |
| | | (3) | (2) | (4) | (1) | (5) | (6) |
| 21 | CBY | 1.64 | 1.28 | 1.72 | 1.13 | N.A. | N.A. |
| | | (3) | (2) | (4) | (1) | | |

*Note:* N.A. = not available.

[a]On the symbols used, see previous tables and the text.

[b]ARMSE = average of the RMSEs across the five horizons (0–1, . . . , 0–5, or 1, . . . , 5). Entries in parentheses represent ranks according to ARMSE (smallest to largest).

tively. The corresponding ranks of BVAR variant B are also high: third, first, and third. The NBER-ASA consensus forecasts are now almost tied for the first rank with the Sims variant B and rank only third for RGNP and fifth for IPD (cf. rows 2, 5, and 7).

8. Table 1.30 sums up the evidence on the comparative accuracy of the several sets of forecasts included in this study, using the longest series of predictions available for each variable. Here again, root mean square errors averaged across the spans serve as the basis for ranking the forecasts, but only the variant A time-series predictions are used. By this criterion, the group forecasts from the NBER-ASA surveys earned ten first and ten second ranks for the twenty-one variables covered. The median individual forecasts ranked first or second six times, third eleven times, and lower four times. Our BVAR model had equal numbers in the first, second, and third ranks (five each), plus six lower ranks. The Michigan (RSQE) forecasts, available for ten variables, ranked first four times, second once, third three times, and fourth twice. The Sims probabilistic model forecasts, available for six variables, were mostly less accurate, and the ARIMA model forecasts were throughout least accurate.

9. Finally, table 1.31, using sums of the ranks across variables, shows that the group (consensus) forecasts from the survey performed best overall in each of the periods covered; the Michigan forecasts were second best; the median individual forecasts, the BVAR model forecasts, and the Sims forecasts share mostly the ranks third or fourth (there are ties); and the ARIMAs rank last. Note that major deviations from this ordering appear for some variables; notably, Michigan is best for UR, Sims for IP. Also, these results con-

**Table 1.31** **Six Sets of Forecasts Ranked According to Their Overall Accuracy across Variables, by Period, 1968:IV–1990:I, 1968:IV–1981:II, and 1981:II–1990:I**

| | | Ranking According to the Sum of Ranks across Variables[b] | | | | | |
|---|---|---|---|---|---|---|---|
| Row | Number of Variables[a] | Individual Median Forecast (1) | Group (Consensus) Forecast (2) | BVAR Model Forecast (3) | Michigan (RSQE) Forecast (4) | Sims Probabilistic Model Forecast (5) | ARIMA Model Forecast (6) |
| | | 1968:IV–1990:I | | | | | |
| 1 | 7 (1–7) | Third (23) | First (13) | Second (18) | N.A. | N.A. | N.A. |
| 2 | 4 (1–3, 6) | Third (14) | First (7) | Fourth (15) | Second (9) | Fourth (15) | Sixth (36) |
| | | 1968:IV–1981:II | | | | | |
| 3 | 4 (8–11) | Second (8) | First (6) | Third (11) | N.A. | N.A. | N.A. |
| | | 1981:II–1990:I | | | | | |
| 4 | 10 (12–21) | Third (30) | First (14) | Second (27) | N.A. | N.A. | N.A. |
| 5 | 5 (14–16, 20, 21) | Third (17) | First (8) | Fourth (18) | Second (11) | N.A. | N.A. |
| 6 | 2 (15, 20) | Fourth (8) | First (3) | Third (6) | Second (5) | Fourth (8) | Sixth (12) |

*Note:* N.A. = not available.
[a]Identified by rows in table 1.30 (in parentheses).
[b]Sum of the ranks from table 1.30 is given in parentheses.

ceal the differences between the forecast horizons, which are sometimes important (e.g., the Michigan forecasts would rate higher for the longer, and lower for the short, spans).

10. It is important to emphasize that these comparisons concentrate on only one aspect of the forecasts and need not imply an overall superiority of any of them. For example, the econometric and time-series models are clearly much better defined, more easily explained, more easily replicated, and more internally consistent than the survey forecasts. But the survey data collectively embody a great deal of apparently useful knowledge and information available to professional forecasters. An interesting project, which must be left for future research, would be to identify the best of the individual forecasts from the surveys and to combine them with each other and with very different model forecasts. Regressions of actual values on predictions from different sources and models would serve as one method for implementing this objective. Given rich data from active forecasters and interesting models, studies of this type should yield useful lessons.

# Appendix
## *Forecasts of Diverse Macroeconomic Series, 1968–81 and 1981–90*

### Selected Nominal Aggregates, 1968–81

Current-dollar expenditures on durable consumer goods, business plant and equipment, and national defense (CD, PE, and DEF, respectively) all contribute strongly to the cyclical nature and volatility of quarterly changes in GNP. DEF is generally treated as an important exogenous variable.

The statistics for mean errors suggest that underestimates prevailed in the forecasts of CD and PE and overestimates in those of DEF, but there is much dispersion across the individual respondents here, which increases strongly with the forecast horizon (table 1A.1, cols. 1–3). The RMSE measures show much the same kind of progression (cols. 4–6). The forecast errors generally are at least smaller than the actual percentage changes in CD, PE, and DEF, but often not by much, as can be seen by comparing the corresponding entries in columns 4–6 and 9–11. Most of the SK and KU values are small, and only a few for CD and DEF may be significant (cols. 7 and 8).

The gains from averaging across the individual forecasts are modest for CD and DEF and, perhaps surprisingly, barely existent for PE (table 1A.2, cols. 1-4). The RMSE ratios for PE are closely clustered, indicating lack of differentiation among the forecasts for this variable. One possible reason may be the availability and influence of the quarterly anticipation series for plant and equipment outlays.

The RMSEs of our BVAR forecasts are larger than those of the group mean forecasts for CD and much larger for DEF. In contrast, BVAR is found to be much more accurate than the survey averages for PE. Indeed, the RMSE ratios i/bv are relatively very low for CD and DEF and very high for PE throughout. (Compare the corresponding measures in cols. 1–4 and 5–8.)

Finally, the NBER-ASA survey questionnaire used through 1981:II asked for forecasts of the levels of inventory investment in current dollars (CBI), another important but highly volatile and hard-to-predict variable. Table 1A.3 shows that the mean errors and root mean square errors for CBI increased markedly with the span while the corresponding standard deviation did not (rows 1–5). The M and MD statistics for mean errors are all positive here; the RMSE for target quarter 5 (i.e., $t + 4$) is about equal to the actual RMSV.

Apparently, CBI is another of those rare cases in which combining the individual survey forecasts is of little help. The group mean's RMSE is relatively large, and even the lower-quartile i/g ratios are close to one (see table 1A.4, rows 1–5). However, our BVAR model performs somewhat worse still here (cf. cols. 1–4 and 5–8).

Table 1A.1 **Selected Measures of Forecast Accuracy and Actual Values: Percentage Changes in Expenditures for Consumer Durable Goods, Plant and Equipment, and National Defense, by Span, 1968–81**

| | | Mean Error | | | Root Mean Square Error | | | | | Actual Value | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Row | Span (Qs) | M (1) | SD (2) | MD (3) | M (4) | SD (5) | MD (6) | SK (7) | KU (8) | M (9) | SD (10) | RMSV (11) |
| | | Consumer Expenditures for Durable Goods (CD) | | | | | | | | | | |
| 1 | 0–1 | −.79 | .73 | −.87 | 3.60 | 1.06 | 3.42 | .88 | .72 | 1.93 | 4.52 | 4.91 |
| 2 | 0–2 | −1.10 | 1.26 | −1.18 | 5.35 | 1.80 | 4.95 | 3.97 | 24.35 | 3.96 | 6.00 | 7.19 |
| 3 | 0–3 | −1.30 | 1.59 | −1.26 | 6.30 | 1.52 | 5.96 | 1.45 | 2.69 | 5.89 | 7.12 | 9.24 |
| 4 | 0–4 | −.97 | 2.07 | −.87 | 6.94 | 1.83 | 6.45 | 1.73 | 4.01 | 7.86 | 7.57 | 10.91 |
| 5 | 0–5 | −1.23 | 2.61 | −.74 | 8.39 | 3.50 | 7.46 | 4.01 | 23.83 | 9.96 | 8.68 | 13.21 |
| | | Plant and Equipment Expenditures (PE) | | | | | | | | | | |
| 6 | 0–1 | −.71 | 1.02 | −.51 | 5.49 | 1.75 | 5.69 | .06 | −.89 | 2.34 | 6.10 | 6.53 |
| 7 | 0–2 | −1.24 | 1.89 | −1.24 | 8.49 | 2.52 | 9.00 | .40 | −.19 | 4.66 | 9.36 | 10.46 |
| 8 | 0–3 | −1.69 | 2.55 | −1.95 | 11.51 | 2.84 | 11.82 | −.07 | −.70 | 6.89 | 11.99 | 13.83 |
| 9 | 0–4 | −1.83 | 3.12 | −1.95 | 13.78 | 3.26 | 14.17 | −.32 | −.46 | 8.86 | 14.19 | 16.73 |
| 10 | 0–5 | −2.77 | 4.18 | −2.46 | 16.46 | 3.68 | 16.79 | −.60 | .01 | 10.83 | 15.97 | 19.30 |
| | | National Defense Expenditures (DEF) | | | | | | | | | | |
| 11 | 0–1 | .16 | .55 | .24 | 2.33 | .56 | 2.25 | .54 | .71 | 1.43 | 2.51 | 2.89 |
| 12 | 0–2 | .21 | .97 | .35 | 3.48 | .97 | 3.41 | .72 | 2.28 | 2.97 | 4.27 | 5.20 |
| 13 | 0–3 | .17 | 1.52 | .40 | 4.19 | 1.21 | 4.05 | 1.02 | 2.88 | 4.75 | 5.98 | 7.64 |
| 14 | 0–4 | .04 | 2.06 | .34 | 4.80 | 1.57 | 4.51 | 1.80 | 6.39 | 6.60 | 7.58 | 10.05 |
| 15 | 0–5 | −.23 | 2.62 | N.A. | 5.79 | 1.94 | 5.38 | 2.21 | 8.93 | 8.60 | 9.55 | 12.85 |

*Note:* On the symbols used, see previous tables and the text.

Table 1A.2 **Individual, Group Mean, and BVAR Forecasts of Percentage Changes in Expenditures for Consumer Durable Goods, Plant and Equipment, and National Defense: Selected Comparisons, by Span, 1968–81**

| Row | Span (Qs) | Group Mean RMSE (1) | RMSE Ratios i/g $Q_1$ (2) | RMSE Ratios i/g MD (3) | RMSE Ratios i/g $Q_3$ (4) | BVAR RMSE (5) | RMSE Ratios i/bv $Q_1$ (6) | RMSE Ratios i/bv MD (7) | RMSE Ratios i/bv $Q_3$ (8) |
|---|---|---|---|---|---|---|---|---|---|
| | | Consumer Expenditures for Durable Goods (CD) | | | | | | | |
| 1 | 0–1 | 3.23 | .95 | 1.12 | 1.34 | 3.98 | .72 | .93 | 1.05 |
| 2 | 0–2 | 4.70 | 1.01 | 1.11 | 1.23 | 5.23 | .88 | 1.02 | 1.15 |
| 3 | 0–3 | 5.38 | 1.02 | 1.13 | 1.31 | 6.27 | .88 | .98 | 1.19 |
| 4 | 0–4 | 5.50 | 1.06 | 1.16 | 1.36 | 6.74 | .84 | .96 | 1.13 |
| 5 | 0–5 | 6.42 | 1.04 | 1.17 | 1.43 | 7.45 | .83 | .99 | 1.18 |
| | | Plant and Equipment Expenditures (PE) | | | | | | | |
| 6 | 0–1 | 5.82 | 1.01 | 1.05 | 1.09 | 2.19 | 1.87 | 2.72 | 3.04 |
| 7 | 0–2 | 8.96 | 1.01 | 1.07 | 1.10 | 3.16 | 2.19 | 2.83 | 3.33 |
| 8 | 0–3 | 11.54 | 1.00 | 1.06 | 1.10 | 4.47 | 2.36 | 2.79 | 3.20 |
| 9 | 0–4 | 13.76 | 1.00 | 1.05 | 1.12 | 5.51 | 2.39 | 2.70 | 3.30 |
| 10 | 0–5 | 15.69 | .98 | 1.05 | 1.13 | 6.41 | 2.50 | 2.89 | 3.55 |
| | | National Defense Expenditures (DEF) | | | | | | | |
| 11 | 0–1 | 1.78 | 1.07 | 1.19 | 1.38 | 2.85 | .66 | .75 | .87 |
| 12 | 0–2 | 2.73 | 1.05 | 1.16 | 1.34 | 5.78 | .47 | .54 | .63 |
| 13 | 0–4 | 3.31 | 1.02 | 1.17 | 1.38 | 8.53 | .38 | .45 | .54 |
| 14 | 0–4 | 3.91 | 1.03 | 1.20 | 1.44 | 10.55 | .35 | .43 | .50 |
| 15 | 0–5 | 4.96 | 1.05 | 1.22 | 1.44 | 13.25 | .32 | .38 | .48 |

*Note:* On the symbols used, see previous tables and the text.

## Components of Real GNP, 1981–90

After mid-1981, the survey collected forecasts of the main GNP expenditure categories in constant dollars. We start with real inventory investment (RCBI), to follow up on the preceding discussion. It turns out that the RCBI forecasts for 1981–90, like the CBI forecasts for 1981–90, have RMSEs that are large relative to the average actual levels and their variability, especially for the more distant target quarters (table 1A.3, rows 6–10). The average MEs are negative but very small, the SDs large and stable. Again, little is gained by averaging the individual forecasts, but the group mean forecasts do have a distinct advantage over the BVAR forecasts (table 1A.4, rows 6–10).

Similarly, real net exports (NX) were on the whole poorly predicted in the 1980s, as seen from the large relative size of the summary error measures in table 1A.3, rows 11–15. For NX, too, the group mean forecasts do not help much, but, in this case, the BVAR forecasts are found to be much more accurate (table 1A.4, rows 11–15).

**Table 1A.3** **Selected Measures of Forecast Accuracy and Actual Values: Nominal and Real Inventory Investment and Real Net Exports, by Span, 1968–81 and 1981–90**

| | | Mean Error | | | Root Mean Square Error | | | | | Actual Value | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Row | Target Quarter | M (1) | SD (2) | MD (3) | M (4) | SD (5) | MD (6) | SK (7) | KU (8) | M (9) | SD (10) | RMSV (11) |
| | | Change in Business Inventories, 1968–81 (CBI) | | | | | | | | | | |
| 1 | 1 | .43 | 2.07 | .45 | 9.33 | 3.65 | 9.76 | .13 | −.26 | 7.30 | 11.27 | 13.43 |
| 2 | 2 | .92 | 2.42 | 1.10 | 9.94 | 3.62 | 10.24 | −.19 | −.49 | | | |
| 3 | 3 | 1.59 | 2.28 | 1.73 | 10.92 | 3.95 | 11.59 | −.46 | −.84 | | | |
| 4 | 4 | 2.57 | 2.53 | 2.78 | 11.63 | 3.77 | 11.84 | −.34 | −.43 | | | |
| 5 | 5 | 3.54 | 2.78 | 3.34 | 23.01 | 3.90 | 13.45 | −.53 | −.21 | | | |
| | | Change in Business Inventories in Constant Dollars, 1981–90 (RCBI) | | | | | | | | | | |
| 6 | 1 | −.20 | 2.94 | −.18 | 18.27 | 4.64 | 18.58 | .58 | 1.35 | 14.58 | 20.37 | 25.05 |
| 7 | 2 | −.47 | 3.57 | .27 | 18.90 | 3.72 | 19.26 | −.02 | −.07 | | | |
| 8 | 3 | −1.09 | 3.26 | −.83 | 19.70 | 3.78 | 19.35 | .47 | .35 | | | |
| 9 | 4 | −.08 | 3.73 | −.64 | 20.46 | 4.10 | 20.25 | .41 | −.22 | | | |
| 10 | 5 | −.69 | 4.41 | −.58 | 21.06 | 4.47 | 21.03 | .70 | .75 | | | |
| | | Net Exports of Goods and Services in Constant Dollars, 1981–90 (NX) | | | | | | | | | | |
| 11 | 1 | 9.33 | 6.02 | 7.48 | 28.83 | 9.87 | 31.32 | −.90 | .35 | −53.19 | 66.09 | 84.84 |
| 12 | 2 | 14.52 | 8.81 | 12.99 | 37.89 | 12.21 | 40.35 | −1.15 | .94 | | | |
| 13 | 3 | 21.36 | 10.72 | 18.35 | 47.60 | 12.79 | 48.36 | −1.02 | 1.21 | | | |
| 14 | 4 | 27.04 | 12.72 | 27.36 | 54.15 | 12.06 | 55.81 | −1.09 | 1.68 | | | |
| 15 | 5 | 31.53 | 15.13 | 28.56 | 60.21 | 12.31 | 60.63 | −1.38 | 2.89 | | | |

*Note:* On the symbols used, see previous tables and the text.

Table 1A.4 **Individual, Group Mean, and BVAR Forecasts of Nominal and Real Inventory Investment and Real Net Exports: Selected Comparisons, by Span, 1968–81 and 1981–90**

| Row | Target Quarter | Group Mean RMSE (1) | RMSE Ratios i/g $Q_1$ (2) | RMSE Ratios i/g MD (3) | RMSE Ratios i/g $Q_3$ (4) | BVAR,[a] RMSE (5) | RMSE Ratios i/bv $Q_1$ (6) | RMSE Ratios i/bv MD (7) | RMSE Ratios i/bv $Q_3$ (8) |
|---|---|---|---|---|---|---|---|---|---|
| | | Change in Business Inventories, 1968–81 (CBI) | | | | | | | |
| 1 | 1 | 10.57 | .98 | 1.07 | 1.15 | 10.70 | .83 | 1.03 | 1.15 |
| 2 | 2 | 10.75 | .97 | 1.07 | 1.18 | 13.32 | .72 | .89 | 1.02 |
| 3 | 3 | 12.69 | .99 | 1.07 | 1.16 | 14.17 | .76 | .95 | 1.07 |
| 4 | 4 | 13.41 | 1.00 | 1.06 | 1.12 | 16.24 | .78 | .91 | 1.09 |
| 5 | 5 | 14.50 | 1.00 | 1.05 | 1.12 | 15.28 | .81 | .99 | 1.12 |
| | | Change in Business Inventories in Constant Dollars, 1981–90 (RCBI) | | | | | | | |
| 6 | 1 | 17.87 | .96 | 1.03 | 1.14 | 19.78 | .76 | .91 | 1.08 |
| 7 | 2 | 18.35 | .99 | 1.06 | 1.12 | 25.95 | .52 | .72 | .82 |
| 8 | 3 | 19.19 | 1.00 | 1.05 | 1.13 | 28.45 | .50 | .66 | .80 |
| 9 | 4 | 19.40 | .97 | 1.06 | 1.20 | 29.38 | .52 | .68 | .95 |
| 10 | 5 | 20.01 | .98 | 1.02 | 1.11 | 29.89 | .55 | .68 | .84 |
| | | Net Exports of Goods and Services in Constant Dollars, 1981–90 (NX) | | | | | | | |
| 11 | 1 | 28.04 | 1.00 | 1.03 | 1.07 | 13.28 | 2.06 | 2.30 | 2.60 |
| 12 | 2 | 36.45 | 1.00 | 1.04 | 1.14 | 17.71 | 2.00 | 2.21 | 2.65 |
| 13 | 3 | 44.91 | .98 | 1.03 | 1.11 | 19.66 | 2.17 | 2.45 | 2.78 |
| 14 | 4 | 52.11 | .98 | 1.03 | 1.10 | 23.08 | 2.16 | 2.36 | 2.51 |
| 15 | 5 | 59.44 | .97 | 1.02 | 1.08 | 31.39 | 1.79 | 1.94 | 2.13 |

*Note:* On the symbols used, see previous tables and the text.

One would expect total consumption (PCE), the largest and smoothest component of real GNP, to be the easiest to predict and in fact the best predicted. A relatively small but smooth and presumably also well-predicted series should be that of state and local government purchases (SLGP). Federal government purchases (FGP) are more autonomous and volatile, hence more difficult to forecast. Residential fixed investment (RFI) is another hard problem for the forecasters, although for different reasons: it is highly cyclical and an early leading series (construction lags behind housing permits and starts are short). Nonresidential fixed investment (NFI) has more persistence, more of an upward trend, and lags at cyclical turning points, which should make it more easily predicted than RFI. Also, NFI is anticipated with long leads by new capital appropriations and contracts and orders for plant and equipment—but these monthly series on business investment commitments are themselves very volatile.

The evidence on the forecasts of percentage changes in PCE, NFI, RFI,

**Table 1A.5 Selected Measures of Forecast Accuracy and Actual Values: Percentage Changes in Consumption, Investment, and Government Components of Real GNP, by Span, 1981–90**

| | | Mean Error | | | Root Mean Square Error | | | | | Actual Value | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Row | Span (Qs) | M (1) | SD (2) | MD (3) | M (4) | SD (5) | MD (6) | SK (7) | KU (8) | M (9) | SD (10) | RMSV (11) |
| | | Personal Consumption Expenditures (PCE) | | | | | | | | | | |
| 1 | 0–1 | −.14 | .20 | −.14 | .83 | .30 | .68 | 1.61 | 2.17 | .78 | .69 | 1.04 |
| 2 | 0–2 | −.26 | .29 | −.29 | 1.10 | .42 | .94 | 1.81 | 2.67 | 1.58 | .99 | 1.86 |
| 3 | 0–3 | −.47 | .46 | −.44 | 1.45 | .51 | 1.27 | 1.96 | 3.74 | 2.44 | 1.22 | 2.73 |
| 4 | 0–4 | −.63 | .65 | −.59 | 1.80 | .64 | 1.60 | 1.96 | 4.33 | 3.33 | 1.41 | 3.62 |
| 5 | 0–5 | −.85 | .85 | −.74 | 2.23 | .86 | 2.00 | 2.03 | 4.02 | 4.22 | 1.69 | 4.55 |
| | | Nonresidential Fixed Investment (NFI) | | | | | | | | | | |
| 6 | 0–1 | −.59 | .91 | −.71 | 2.85 | 2.10 | 2.43 | 5.07 | 26.62 | 1.05 | 2.00 | 2.80 |
| 7 | 0–2 | −1.22 | 1.42 | −1.26 | 4.53 | 1.91 | 4.07 | 3.99 | 18.39 | 2.21 | 4.40 | 4.92 |
| 8 | 0–3 | −1.47 | 1.88 | −1.51 | 6.76 | 2.51 | 5.86 | 2.32 | 5.35 | 3.46 | 6.18 | 7.08 |
| 9 | 0–4 | −2.31 | 2.38 | −2.36 | 8.47 | 2.14 | 7.97 | 1.43 | 1.69 | 4.87 | 7.85 | 9.24 |
| 10 | 0–5 | −2.76 | 3.14 | −2.97 | 10.33 | 2.55 | 9.99 | 1.35 | 1.81 | 6.46 | 9.26 | 11.29 |
| | | Residential Fixed Investment (RFI) | | | | | | | | | | |
| 11 | 0–1 | −.85 | 1.08 | −.67 | 4.16 | 1.52 | 3.83 | 1.89 | 4.70 | 1.18 | 4.84 | 4.98 |
| 12 | 0–2 | −1.87 | 2.27 | −1.42 | 7.56 | 2.48 | 7.18 | .67 | −.24 | 2.95 | 9.01 | 4.27 |
| 13 | 0–3 | −2.93 | 3.19 | −2.32 | 10.23 | 3.28 | 10.14 | .36 | −.21 | 5.06 | 12.59 | 13.57 |
| 14 | 0–4 | −4.12 | 3.93 | −3.56 | 12.68 | 4.26 | 12.42 | .28 | −.67 | 7.39 | 15.78 | 17.42 |
| 15 | 0–5 | −5.58 | 5.23 | −4.94 | 14.95 | 5.07 | 14.40 | .21 | −.95 | 9.90 | 18.71 | 21.17 |
| | | Federal Government Purchases (FGP) | | | | | | | | | | |
| 16 | 0–1 | −.60 | 1.33 | −.51 | 3.99 | .81 | 3.77 | .60 | −.37 | 1.16 | 4.15 | 4.31 |
| 17 | 0–2 | −.79 | 1.50 | −.90 | 5.25 | 1.32 | 5.03 | 1.49 | 3.63 | 2.22 | 5.24 | 5.69 |
| 18 | 0–3 | −.94 | 1.75 | −1.21 | 5.35 | 1.43 | 5.01 | 1.78 | 4.47 | 3.09 | 5.28 | 6.12 |
| 19 | 0–4 | −.74 | 2.36 | −1.35 | 5.91 | 3.06 | 5.27 | 3.75 | 17.22 | 4.00 | 4.85 | 6.29 |
| 20 | 0–5 | −1.55 | 2.70 | −1.81 | 6.16 | 1.71 | 5.74 | 1.14 | 1.48 | 5.27 | 6.05 | 8.02 |
| | | State and Local Government Purchases (SLGP) | | | | | | | | | | |
| 21 | 0–1 | −.13 | .28 | −.17 | .90 | .33 | .85 | 1.87 | 4.40 | .52 | .70 | .87 |
| 22 | 0–2 | −.24 | .52 | −.26 | 1.24 | .56 | 1.15 | 2.77 | 9.44 | 1.12 | 1.07 | 1.55 |
| 23 | 0–3 | −.38 | .72 | −.40 | 1.57 | .77 | 1.52 | 2.61 | 9.42 | 1.72 | 1.32 | 2.17 |
| 24 | 0–4 | −.62 | .88 | −.58 | 1.89 | .92 | 1.79 | 2.98 | 12.01 | 2.34 | 1.54 | 2.80 |
| 25 | 0–5 | −.92 | 1.09 | −1.13 | 2.35 | 1.11 | 2.03 | 2.82 | 11.02 | 2.99 | 1.85 | 3.52 |

*Note:* On the symbols used, see previous tables and the text.

FGP, and SLGP is generally consistent with these priors. Thus, forecasts of growth in PCE four quarters ahead have errors averaging about half the actual percentage change (table 1A.5, rows 1–5). This is not great, but fair, and in sharp contrast to the apparent failure of forecasts of inventory investment (the least predictable of the components of aggregate demand). The RMSEs of the NFI forecasts are much smaller than their counterparts for RFI (but the actual percentage changes are also smaller for NFI; compare the corresponding entries in rows 6–10 and 11–15). The SLGP forecasts are definitely much more accurate than the FGP forecasts (cf. rows 16–20 and 21–25).

The forecasts share some characteristics across all the variables. All the M and MD statistics for mean errors are negative, suggesting a prevalence of underprediction errors (cols. 1 and 3). The absolute values of these statistics increase with the span in each case. Indeed, all the summary error measures, except SK and KU, show such increases, as do the statistics for actual values.[17] The means of the RMSEs are generally larger than the medians, and SK is greater than zero. The KU statistics are large in some cases, particularly for NFI (short forecasts) and SLGP.

Combining the individual forecasts into group means reduces the RMSEs for each variable and span, as can be seen by comparing column 1 of table 1A.6 with column 6 (and a fortiori with col. 4) of table 1A.5. At the lower quartile $Q_1$, the RMSE ratios i/g are close to one throughout; the range of the median ratios is about 1.1–1.3, and that of the $Q_3$ ratios is 1.2–1.7. The group mean forecasts perform best (the ratios are highest) for PCE and SLGP (see table 1A.6, cols. 2–4).

Our BVAR forecasts have larger RMSEs than the NBER-ASA group mean forecasts 80 percent of the time, according to the paired entries in columns 1 and 5 of table 1A.6. They are very poor for PCE and definitely inferior for FGP, whereas elsewhere the differences are much smaller (cf. cols. 1–4 and 5–8).

### Consumer Price Inflation and Interest Rates, 1981–90

Forecasters underpredicted CPI inflation just as they did IPD inflation (see the negative signs of the mean errors in table 1A.7, rows 1–5, cols. 1 and 3). The RMSEs of these forecasts are discouragingly large compared to the descriptive statistics for the actual values (cf. cols. 4 and 6 with 11, in particular). Note that the NBER-ASA survey questionionnaire asked directly for forecasts of the level of *CPI inflation* at an annual rate in the current quarter and the following four quarters (not for forecasts of the CPI itself).

In contrast, the forecasts of the three-month Treasury-bill rate (TBR) had relatively small errors according to these comparisons (rows 6–10). The fore-

17. A few deviations from the rule appear in the longest forecasts; they are apparently due to outliers and small-sample problems.

able 1A.6 **Individual, Group Mean, and BVAR Forecasts of Percentage Changes in Consumption, Investment, and Government Components of Real GNP: Selected Comparisons, by Span, 1981–90**

| | | | RMSE Ratios i/g | | | | RMSE Ratios i/bv | | |
|---|---|---|---|---|---|---|---|---|---|
| ow | Span (Qs) | Group Mean RMSE (1) | $Q_1$ (2) | MD (3) | $Q_3$ (4) | BVAR RMSE (5) | $Q_1$ (6) | MD (7) | $Q_3$ (8) |
| | | Personal Consumption Expenditures (PCE) | | | | | | | |
| 1 | 0–1 | .58 | 1.08 | 1.23 | 1.73 | 1.97 | .31 | .36 | .48 |
| 2 | 0–2 | .79 | 1.12 | 1.22 | 1.41 | 3.78 | .22 | .24 | .28 |
| 3 | 0–3 | .98 | 1.11 | 1.25 | 1.51 | 5.65 | .20 | .22 | .27 |
| 4 | 0–4 | 1.18 | 1.15 | 1.24 | 1.71 | 7.66 | .18 | .20 | .25 |
| 5 | 0–5 | 1.47 | 1.10 | 1.27 | 1.66 | 9.37 | .18 | .21 | .24 |
| | | Nonresidential Fixed Investment (NFI) | | | | | | | |
| 6 | 0–1 | 2.01 | 1.10 | 1.19 | 1.34 | 2.23 | .90 | 1.02 | 1.19 |
| 7 | 0–2 | 3.43 | 1.03 | 1.11 | 1.26 | 3.64 | .95 | 1.04 | 1.17 |
| 8 | 0–3 | 4.99 | 1.04 | 1.12 | 1.26 | 5.31 | .96 | 1.04 | 1.19 |
| 9 | 0–4 | 6.89 | 1.00 | 1.06 | 1.22 | 6.83 | .99 | 1.09 | 1.23 |
| 0 | 0–5 | 8.69 | .99 | 1.05 | 1.18 | 8.25 | 1.02 | 1.09 | 1.26 |
| | | Residential Fixed Investment (RFI) | | | | | | | |
| 1 | 0–1 | 3.01 | 1.05 | 1.26 | 1.45 | 3.97 | .79 | .97 | 1.21 |
| 2 | 0–2 | 5.83 | .96 | 1.25 | 1.43 | 5.24 | 1.10 | 1.30 | 1.80 |
| 3 | 0–3 | 8.42 | .94 | 1.21 | 1.41 | 8.33 | .90 | 1.15 | 1.51 |
| 4 | 0–4 | 10.63 | .93 | 1.24 | 1.42 | 11.30 | .80 | 1.02 | 1.35 |
| 5 | 0–5 | 12.62 | .96 | 1.15 | 1.45 | 14.64 | .71 | .98 | 1.23 |
| | | Federal Government Purchases (FGP) | | | | | | | |
| 6 | 0–1 | 3.31 | 1.00 | 1.14 | 1.27 | 4.61 | .72 | .80 | .95 |
| 7 | 0–2 | 4.22 | 1.02 | 1.10 | 1.29 | 7.26 | .59 | .65 | .82 |
| 8 | 0–3 | 4.11 | 1.04 | 1.18 | 1.37 | 9.02 | .50 | .54 | .66 |
| 9 | 0–4 | 3.79 | 1.13 | 1.31 | 1.48 | 10.39 | .39 | .52 | .63 |
| 0 | 0–5 | 4.55 | 1.07 | 1.30 | 1.48 | 11.44 | .44 | .53 | .65 |
| | | State and Local Government Purchases (SLGP) | | | | | | | |
| 1 | 0–1 | .61 | 1.14 | 1.23 | 1.51 | .49 | 1.40 | 1.64 | 2.07 |
| 2 | 0–2 | .75 | 1.02 | 1.19 | 1.39 | .82 | 1.09 | 1.44 | 1.62 |
| 3 | 0–3 | .91 | 1.00 | 1.17 | 1.27 | 1.20 | 1.11 | 1.41 | 1.60 |
| 4 | 0–4 | 1.07 | 1.13 | 1.19 | 1.40 | 1.68 | 1.13 | 1.29 | 1.54 |
| 5 | 0–5 | 1.35 | 1.01 | 1.13 | 1.30 | 2.18 | 1.04 | 1.28 | 1.53 |

*ote:* On the symbols used, see previous tables and the text.

**Table 1A.7** **Selected Measures of Forecast Accuracy and Actual Values: Consumer Price Inflation, the Treasury-Bill Rate, and the Corporate Bond Yield, 1981–90**

| | | Mean Error | | | Root Mean Square Error | | | | | Actual Value | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Row | Target Quarter | M (1) | SD (2) | MD (3) | M (4) | SD (5) | MD (6) | SK (7) | KU (8) | M (9) | SD (10) | RMSV (11) |
| | | Consumer Price Index, % Change (CPI) | | | | | | | | | | |
| 1 | 1 | −.61 | .47 | −.50 | 1.20 | .39 | 1.08 | .34 | −1.37 | 1.02 | .54 | 1.16 |
| 2 | 2 | −.58 | .48 | −.46 | 1.18 | .43 | 1.15 | .11 | −1.41 | | | |
| 3 | 3 | −.57 | .50 | −.46 | 1.20 | .42 | 1.20 | .19 | −1.36 | | | |
| 4 | 4 | −.53 | .50 | −.51 | 1.21 | .38 | 1.26 | .10 | −1.50 | | | |
| 5 | 5 | −.51 | .53 | −.50 | 1.23 | .39 | 1.24 | .00 | −1.41 | | | |
| | | Treasury-Bill Rate, 3-Month, % (TBR) | | | | | | | | | | |
| 6 | 1 | .03 | .24 | −.03 | .49 | .36 | .40 | 2.50 | 8.01 | 8.38 | 2.48 | 8.74 |
| 7 | 2 | .19 | .49 | .24 | 1.51 | .51 | 1.52 | .06 | 2.75 | | | |
| 8 | 3 | .39 | .68 | .61 | 1.85 | .50 | 1.80 | .11 | .85 | | | |
| 9 | 4 | .65 | .59 | .83 | 2.03 | .56 | 2.23 | −1.55 | 2.14 | | | |
| 10 | 5 | 1.05 | .68 | 1.03 | 2.48 | .71 | 2.62 | −.42 | 1.09 | | | |
| | | Corporate Bond Yield, % (CBY) | | | | | | | | | | |
| 11 | 1 | −.26 | .30 | −.24 | .72 | .35 | .63 | 2.01 | 5.65 | 11.64 | 2.34 | 11.87 |
| 12 | 2 | −.00 | .48 | .12 | 1.59 | .59 | 1.37 | .66 | −.26 | | | |
| 13 | 3 | .21 | .61 | .31 | 1.86 | .49 | 1.81 | .78 | .51 | | | |
| 14 | 4 | .39 | .72 | .58 | 2.05 | .43 | 2.02 | .39 | −.28 | | | |
| 15 | 5 | .60 | .74 | .69 | 2.25 | .66 | 2.36 | .88 | 1.58 | | | |

*Note:* On the symbols used, see previous tables and the text.

**Table 1A.8** **Individual, Group Mean, and BVAR Forecasts of Consumer Price Inflation, the Treasury-Bill Rate, and the Corporate Bond Yield: Selected Comparisons, by Span, 1981–90**

| | | | RMSE Ratios i/g | | | | RMSE Ratios i/bv | | |
|---|---|---|---|---|---|---|---|---|---|
| Row | Target Quarter | Group Mean RMSE (1) | $Q_1$ (2) | MD (3) | $Q_3$ (4) | BVAR,[a] RMSE (5) | $Q_1$ (6) | MD (7) | $Q_3$ (8) |
| | | Consumer Price Index, % Change (CPI) | | | | | | | |
| 1 | 1 | .53 | 1.01 | 1.03 | 1.12 | .54 | 1.63 | 2.63 | 3.22 |
| 2 | 2 | .46 | 1.01 | 1.03 | 1.12 | .74 | 1.02 | 1.63 | 2.29 |
| 3 | 3 | .48 | 1.01 | 1.03 | 1.18 | .78 | 1.05 | 1.80 | 2.06 |
| 4 | 4 | .54 | 1.01 | 1.02 | 1.09 | .80 | 1.07 | 1.63 | 1.79 |
| 5 | 5 | .58 | 1.01 | 1.02 | 1.08 | .95 | .92 | 1.36 | 1.59 |
| | | Treasury-Bill Rate, 3-month, % (TBR) | | | | | | | |
| 6 | 1 | .20 | 1.45 | 1.83 | 2.46 | .96 | .29 | .38 | .56 |
| 7 | 2 | .90 | 1.09 | 1.26 | 2.11 | 1.62 | .59 | .81 | 1.17 |
| 8 | 3 | 1.38 | 1.02 | 1.31 | 1.54 | 2.03 | .65 | .80 | 1.08 |
| 9 | 4 | 1.77 | 1.00 | 1.14 | 1.29 | 2.51 | .68 | .77 | .94 |
| 10 | 5 | 2.49 | 1.01 | 1.08 | 1.18 | 3.03 | .67 | .78 | .95 |
| | | Corporate Bond Yield, % (CBY) | | | | | | | |
| 11 | 1 | .38 | 1.17 | 1.57 | 1.81 | .77 | .56 | .77 | .98 |
| 12 | 2 | .83 | 1.15 | 1.48 | 2.52 | 1.26 | .72 | .95 | 1.75 |
| 13 | 3 | 1.24 | 1.09 | 1.25 | 1.74 | 1.74 | .77 | .92 | 1.25 |
| 14 | 4 | 1.51 | 1.03 | 1.19 | 1.46 | 2.19 | .69 | .88 | 1.03 |
| 15 | 5 | 2.42 | .98 | 1.11 | 1.20 | 2.65 | .60 | .83 | 1.02 |

*Note:* On the symbols used, see previous tables and the text.

casts of the (new high-grade) corporate bond yield (CBY) were even more accurate (rows 11–15).

Despite the already noted weakness of most of the individual CPI forecasts, the corresponding group mean forecasts perform relatively well. Their RMSEs are considerably smaller than those of the BVAR model and less than half those of the average individual forecasts (cf. table 1A.8, rows 1–5, and table 1A.7, rows 1–5, cols. 4 and 6). The i/g ratios cluster close to one between $Q_1$ and $Q_2$, which indicates that the forecasts concerned are remarkably alike.

For the interest rates TBR and CBY, combining the individual forecasts greatly reduces errors, but with notable exceptions at the most distant target quarter (5). Here the RMSEs are much larger for the BVAR than the group mean forecasts, and, correspondingly, the i/bv ratios are much lower than the i/g ratios (cf. table 1A.8, rows 6–15, cols. 1–4 and 5–8).

## References

Braun, P. A., and I. Yaniv. 1991. A case study of expert judgment: Economists' probabilistic forecasts. University of Chicago. Mimeo.

Burns, A. F., and W. C. Mitchell. 1946. *Measuring business cycles.* New York: NBER.

Butler, W. F., and R. A. Kavesh. 1974. Judgmental forecasting of the gross national product. In *Methods and techniques of business forecasting,* ed. W. F. Butler, R. A. Kavesh, and R. B. Platt. Englewood Cliffs, N.J.: Prentice-Hall.

Carlson, J. A. 1977. Study of price forecasts. *Annals of Economic and Social Measurement* 6 (1): 27–56.

Clemen, R. T. 1989. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* 5 (4): 559–83.

Cole, R. 1969. *Errors in provisional estimates of gross national product.* New York: Columbia University Press (for the NBER).

Cox, G. V. 1929. *An appraisal for American business forecasts.* Chicago: University of Chicago Press.

Dominguez, K. M., R. C. Fair, and M. D. Shapiro, 1988. Forecasting the Depression: Harvard versus Yale. *American Economic Review* 78 (4): 595–612.

Evans, M. K., Y. Haitovsky, and G. I. Treyz, assisted by V. Su. 1972. An analysis of the forecasting properties of U.S. econometric models. In *Econometric models of cyclical behavior,* ed. B. G. Hickman, 2 vols. New York: Columbia University Press (for the NBER).

Fels, R., and C. E. Hinshaw. 1968. *Forecasting and recognizing business cycle turning points.* New York: Columbia University Press (for the NBER).

Grunberg, E., and F. Modigliani. 1954. The predictability of social events. *Journal of Political Economy* 62 (6): 465–78.

Haitovsky, Y., G. I. Treyz, and V. Su. 1974. *Forecasts with quarterly macroeconometric models.* New York: Columbia University Press (for the NBER).

Hicks, J. R. 1979. *Causality in economics.* New York: Basic.

Keynes, J. M. 1936. *The general theory of employment, interest and money.* London: Macmillan.

Knight, F. M. 1921. *Risk, uncertainty, and profit.* Boston: Houghton Mifflin.

Litterman, R. B. 1986. Forecasting with Bayesian vector autoregressions—five years of experience. *Journal of Business and Economic Statistics* 4 (1): 25–38.

McCloskey, D. N. 1988. The limits of expertise: If you're so smart, why ain't you rich? *American Scholar* (Summer), 393–406.

McNees, S. K. 1979. The forecasting record of the 1970s. *New England Economic Review* (September/October) 57(3): 1–21.

Mincer, J. 1969. *Economic forecasts and expectations: Analyses of forecasting behavior and performance.* New York: Columbia University Press (for the NBER).

Moore, G. H. 1969. Forecasting short-term economic change. *Journal of the American Statistical Association* 64 (March): 1–22.

Morgenstern, O. 1928. *Wirtschaftsprognose: Eine Untersuchung ihrer Voraussetzungen and Möglichkeiten.* Vienna: Springer.

Nelson, C. R. 1984. A benchmark for the accuracy of econometric forecasts of GNP. *Business Economics* 19 (3): 52–58.

Niemira, M., and P. A. Klein. 1991. Forecasting the Great Depression using cyclical indicators: Another look. New York: Mitsubishi Bank. Mimeo.

Okun, A. M. 1959. A review of some economic forecasts for 1955–57. *Journal of Business* 32 (July): 199–211.

Sims, C. A. 1989. A nine-variable probabilistic macroeconomic forecasting model. Discussion Paper no. 14. Institute for Empirical Macroeconomics. Minneapolis: Federal Reserve Bank of Minnesota, May. Mimeo.

Sims, C. A., and R. M. Todd. 1991. Evaluating Bayesian vector autoregressive forecasting procedures for macroeconomic data: Part 2. Paper presented at the NSF-NBER Seminar on Bayesian Inference in Econometrics and Statistics, St. Paul, Minnesota, April.

Solow, R. M. 1985. Economic history and economics. *American Economic Review* 75 (2): 328–31.

Stekler, H. O. 1968. Forecasting with econometric models: An evaluation. *Econometrica* 36: 437–63.

Su, V., and J. Su. 1975. An evaluation of the ASA/NBER business outlook survey forecasts. *Explorations in Economic Research* 2: 588–618.

Suits, D. B. 1962. Forecasting and analysis with an econometric model. *American Economic Review* 52:104–32.

Theil, H. 1961. *Economic forecasts and policy.* 2d rev. ed. Amsterdam: North-Holland.

———. 1966. *Applied economic forecasting.* Amsterdam: North-Holland; Chicago: Rand McNally.

———. 1967. *Economics and information theory.* Amsterdam: North-Holland; Chicago: Rand McNally.

Thorp, W. L. 1926. *Business annals.* New York: NBER.

Zarnowitz, V. 1967. *An appraisal of short-term economic forecasts.* New York: Columbia University Press (for the NBER).

———. 1968. The new ASA-NBER survey of forecasts by economic statisticians. *American Statistician* 23 (1): 12–16.

———. 1971. New plans and results of research in economic forecasting. In *Fifty-first annual report.* New York: NBER.

———. 1972. Forecasting economic conditions: The record and the prospect. In *The business cycle today,* ed. V. Zarnowitz. New York: Columbia University Press (for the NBER).

———. 1974. How accurate have the forecasts been? In *Methods and techniques of business forecasting,* ed. W. F. Butler, R. A. Kavesh, and R. B. Platt. Englewood Cliffs, N.J.: Prentice-Hall.

———. 1979. An analysis of annual and multiperiod quarterly forecasts of aggregate income, output, and the price level. *Journal of Business* 52 (1): 1–33.

———. 1982. Improvements and new studies of the economic outlook survey. *NBER Reporter* (Summer), 11–13.

———. 1983. Some lessons from research in macroeconomic forecasting. *Economic Outlook USA* 10 (4): 83–86.

———. 1984. The accuracy of individual and group forecasts from business outlook surveys. *Journal of Forecasting* 3 (1): 11–26.

———. 1985. Rational expectations and macroeconomic forecasts. *Journal of Business and Economic Statistics* (4): 293–311.

———. 1991. Has macro-forecasting failed? Working Paper no. 3867. Cambridge, Mass.: NBER, October. (This paper will be reprinted in the proceedings of the Cato Institute's ninth annual monetary conference: *Money, macroeconomics, and forecasting.* Washington, D.C.: Cato Institute, in press.)

———. 1992. *Business cycles: Theories, history, indicators, and forecasting.* Chicago: University of Chicago Press.

Zarnowitz, V., and P. Braun. 1990. Major macroeconomic variables and leading in-

dexes: Some estimates of their interrelations. In *Analyzing modern business cycles: Essays honoring Geoffrey H. Moore,* ed. P. A. Klein. Armonk, N.Y.: M. E. Sharpe.
Zarnowitz, V., and L. A. Lambros. 1987. Consensus and uncertainty in economic prediction. *Journal of Political Economy* 95 (3): 591–621.

# Comment Allen Sinai

Victor Zarnowitz and Phillip Braun (ZB) have produced an impressive paper on the record of macroeconomic forecasts and forecasting performance, probably the most encyclopedic ever on this topic. The data set, while not without deficiencies, is rich in content, comprehensive, and spans an unusually long period, 1968–90. There is a representative enough sample of forecasters and a long enough time frame to promote confidence in the implications for forecasts and forecasting performance suggested by the work.

The macro variables covered were big and small (e.g., ranging from gross national product [GNP] to the change in business inventories), important and less so in the macro scheme of things, showed both little and considerable variability, and presented a wide range of ease or difficulty in forecasting. The forecasters sampled and forecasting methods used represent a reasonable cross section of those who engage in macro forecasting, although classification of the types of forecasts and forecasters is difficult because of the mix in background and methods of the forecasters.

One problem with the data was considerable variability in the survey respondents, over time and across individuals. On average, the response rate was not very high and was quite variable. The number of respondents shrank dramatically in the 1980s compared with the 1970s. There was no systematic statistically based process for generating the observations used to calculate the results. Inferences drawn from so nonrandom a sample, especially one on which distributional summary statistics are calculated, may be more suggestive than confirmational and should be interpreted cautiously.

ZB nearly exhaust the possible range of questions about macro forecasts and macro forecasting, providing numerous insights on forecasting accuracy and performance that can help consumers of economic forecasts in understanding what to expect from forecasts and forecasters.

This Comment discusses several aspects of the ZB paper, drawing out what might be of interest to the historian of forecast performance, to the forecasting practitioner, and to those who are consumers of macroeconomic forecasts. Of interest is what the forecasts of macroeconomic variables reveal about forecast accuracy, what forecasters can learn from past forecasting performance to help

Allen Sinai is chief economist and president, Lehman Brothers Economic Advisors, and adjunct professor, Lemberg Program for International Economics and Finance, Brandeis University.

forecasting in the future, and what consumers of macroeconomic forecasts can reasonably expect from forecasts and forecasters. A couple of specific criticisms are separately treated.

## Forecasts and Forecasting Performance

The ZB paper provides the most comprehensive and exhaustive survey of macroeconomic forecasts presented to date.[1] The sample period covered twenty-two years with eleven variables from 1968:IV to 1981:II and seventeen through 1990:I, included a large number of forecasters, contained a variety of forecasting methods, involved numerous organizational sites from which the forecasts were made, and provided a varied, often volatile set of macroeconomic conditions and events to forecast. Topics covered relative to forecasts and forecasting performance included accuracy (for specific variables and in general), the variability of forecasts, near-term versus longer-run forecasts, "consensus" versus individual forecasts, and forecast performance (by organization, by period, and by method).

Some deficiencies of the data must be noted, however. How representative the sample of forecasts and forecasters is affects the validity of any inferences drawn.

First, the sample generating the forecast observations was not of uniform size and composition in a cross-sectional or time-series sense. Nor was it randomly selected by any sampling method. There were 86 respondents used in the period 1968:IV–1981:2. The mean number per survey, 40.8, was less than half the total. The standard deviation was 11.3, or 27.6 percent of the mean for this subperiod, an indication of considerable variability in the responses across surveys. The sample from 1981:III to 1990:I contained 29 individuals, only one-third the number in the prior subperiod, although the responses were more consistent. The mean number of respondents per survey was 17.2 and the standard deviation 3.3.

Respondent participation was not strong, with only 24.2 surveys per respondent, on average, out of 51 in 1968:IV–1981:II and a standard deviation of 10.4, or 43 percent of the mean. The mean number of surveys per respondent was 20.8 out of 35 but with a smaller standard deviation of 7.5, or 36.1 percent of the mean.

1. McNees has summarized and presented the results of macroeconomic forecasts in numerous articles (e.g., McNees 1979, 1981, 1988; McNees and Ries 1983), but his sample generally has been limited to econometrics-based forecasts. Victor Zarnowitz, and now Phillip Braun, has been collecting a large number of macroeconomic variable forecasts for the North-Holland publication *Economic Forecasts*. Numerous forecasters are surveyed whose methods of forecasting vary considerably. This effort is fairly recent, but the data set will provide an excellent basis for analysis at some point. The *Eggert Blue Chip Survey* collects forecasts from about fifty-three forecasters who are based at different types of organizations and who use different methods to forecast. This could provide data useful for analyzing forecasts and forecasting performance if it would ever be made available for that purpose. The *Wall Street Journal* surveys fifty or so forecasters twice a year, publishing the results in early January and in early July, and has been doing so regularly since 1981. This is yet another source of data on macroeconomic forecasts and forecasting performance.

A smaller sample of participants can produce greater variability in the various summary measures of forecast accuracy, those for central tendency or spread. A wide dispersion of forecasts would have a similar effect. How to correct for the variability of survey responses across time and of respondents by survey is not clear. Means and standard deviations of the individual and group mean forecasts are likely biased in such a situation.

Second, certain differences between forecasters were not accounted for, nor were any adjustments made for the time spent in forecasting. Those who participated in a few surveys might not have been similar in interest to those who responded to many surveys or have had similar training, experience, or ability. "Part-time" forecasting, on average, probably produces less accuracy, "full-time" forecasting more, and "commercially based" forecasting perhaps the most accuracy.

Third, coverage varied across individuals and over time, with turnover of participants high, again a problem for the drawing of inferences from the calculated results.

Fourth, since the sample participants were largely from business but spread across other areas including finance, research and consulting, academic institutions, government, and others, skill levels, possible influences of the work environment on forecasting, and methods used could have been very different. Forecast comparisons need to correct for these factors.

Fifth, the forecasting method most often identified was "informal judgment." This provided another sort of heterogeneity in the sample since the informal judgment method of forecasting is amorphously defined and for at least some forecasters potentially a surrogate for an otherwise identifiable forecast method.

Such features of the data probably bias the results more than if each forecaster exhibited a similar configuration of characteristics and the sample size were the same over time. Doubt is cast on the validity of conclusions drawn because the sample may not have been representative of the population. The inferences claimed from the study must therefore be regarded as tentative and in need of further investigation.

Any weaknesses of the data did not carry over into the assessment of forecasting performance, however. ZB provide a complete and comprehensive set of results and make a major contribution to the issue of forecast accuracy, producing numerous insights and implications for practitioners and users of macroeconomic forecasts.

A full range of summary measures of forecast performance is provided, for both short- and longer-run forecasts. Almost everything is there—mean errors; marginal errors; mean absolute errors (MAEs); root mean square errors (RMSEs); measures of dispersion such as standard deviation, median, interquartile range, skewedness, and kurtosis; early quarter and late quarter forecasts; group mean as consensus versus individual forecasts; levels and percentage changes; zero- to five-quarters-ahead forecasts; Bayesian vector autoregression (BVAR) forecasts; subperiod forecast assessment (1968:IV–1979:III and 1979:IV–1990:I); Michigan (RSQE) model econometric forecasts versus those from the NBER-ASA survey; Sims model (BVAR) forecasts versus the NBER-ASA versus the Sims-Todd ARIMA—in thirty-one tables of data and eight appendix tables, quite a collection of results!

The analysis is also very complete, ranging over issues of absolute accuracy in forecasts of changes and levels; the effects of data revision on accuracy; the distribution of errors in the short run (current to one or two quarters out) and in the long run (as much as five quarters out); accuracy for major variables such as nominal GNP, real GNP, and inflation but also for other macroeconomic variables such as the unemployment rate, housing starts, capital goods spending, net exports, and certain interest rates; individual versus "consensus" or group mean forecasts; how well certain forecasters forecast certain variables; a comparison of forecasts before and after the big change of Federal Reserve policy in October 1979; econometric model performance of the Michigan model versus the NBER-ASA; the NBER-ASA versus time-series methods such as BVAR, the Sims version of BVAR, and ARIMA.

Many of the conclusions drawn have been put forward before, for example, larger forecast errors the longer the forecast horizon and the superior performance of group mean or consensus forecasts relative to most individual forecasts. The supporting evidence is more substantial in ZB than in other studies, however.

Some results are new and interesting. There is new evidence particularly on the performance of BVAR and other time-series forecasting methods versus the NBER-ASA forecasts, which were mostly informal judgment. These relatively new methods of forecasting have been subjected to little evaluation, in terms of accuracy, relative to others such as informal judgment, econometrics, or the leading economic indicators barometric approach. The results from the BVAR approach of ZB were mixed, where two variants were used, one with an assumption of no current-quarter knowledge (variant A) and the other assuming perfect foresight in the quarter (variant B). But the results appear promising enough to warrant additional examination of BVAR as an alternative or supplemental macroeconomic forecasting tool.

Other time-series methods, represented in models developed by Christopher Sims and by Sims and R. Todd, BVAR- and ARIMA-type specifications, were analyzed for the major variables. Some ARIMA forecasts of Charles Nelson's and Frederick Joutz's were also evaluated.

Here, the results were again mixed, with forecasts of some variables showing up relatively well against the NBER-ASA surveys and some not, depending especially on whether the assumption was no current-quarter knowledge or perfect foresight. If the former, the ARIMA comparisons were unfavorable. If the latter, the ARIMA comparisons were competitive. Consensus or group mean forecasts almost uniformly were superior to the ARIMA methods, regardless of which type. The Sims BVAR results were competitive, suggesting BVAR as a forecasting technique worth using, if only at least on a collateral basis.

Group mean forecasts consistently did better than most individual forecasts and the various time-series methods of forecasting, not a new result, but one underscored and most impressive in ZB for the degree and widespread nature of the supporting evidence. This is a significant and important finding for both practitioners and users of macroeconomic forecasts.

A few interesting topics were not analyzed. There was no discussion or evaluation of forecast accuracy on cyclical turns. ZB did not deal with the question of forecasting turning points in major variables. It would have been of interest to see how the NBER-ASA forecasts did on turning points and in identifying the extent and length of downturns and expansions and to see how consistent forecasts were across different cyclical episodes. The 1970s and 1980s were quite volatile, in a business-cycle sense, with both endogenous swings and external shocks, policy and other, perturbing the economy more than usual. How macro forecasts did under such conditions could have been highlighted given the data available, would have been instructive on issues of accuracy, and would have been of interest to practitioners of forecasting.

A useful historical analysis would have carefully examined the forecasts in periods after "shocks" such as the fourfold oil price increase of 1973–74 and the radical change of Federal Reserve policy in October 1979.

A notion of rational expectations is that forecasts, especially those of econometric models and implicitly others that use a structural approach, might not do well after a change in structure. Looking at forecast performance for a few years before and after the structural changes that occurred and across forecasting methods such as econometric, informal judgment, the barometric approach of the leading economic indicators, and anticipatory surveys could have provided a test of this rational expectations idea.

A historical description of how the forecasts fared over business-cycle episodes, à la McNees (1988) and McNees and Ries (1983), would also have been of interest. This cannot be extracted from the tables, although it certainly exists in the ZB data set.

Interesting also was the attempt at a before-and-after comparison of forecast accuracy where the new Fed policy (NFP) under Federal Reserve chairman Volcker constituted the dividing line between 1968:IV–1979:III and 1979:IV–1990:I. Regardless of the forecasting method, a worse performance might have been expected in the period 1979:IV–1990:I than in 1968:IV–1979:III. The structure of the economy certainly shifted during the late 1970s and early 1980s, at first because of oil price shocks and then because of the NFP.

The period comparisons, eleven years each, were interpreted by ZB as not showing any significant improvement or worsening of forecasts between them. But some of the ZB data do not support this assertion. Table 1.14 shows that, in the period 1979–90, the individual forecasts were worse on nominal GNP, worse part of the time on real GNP, and better on inflation. The group mean forecasts were significantly worse on nominal GNP, slightly worse on real GNP near term, and distinctly better on inflation.

In any case, disaggregation over time, rather than the lumping together of so much experience into eleven-year subsamples, would be necessary if differences were to be assessed fully. Macro forecasting is a lot easier in expansion (e.g., the years 1983–88) than over a shorter period of time such as 1981–82, which was characterized by a great deal of volatility.

There is considerable averaging in the two eleven-year spans, perhaps wiping out or diminishing what might appear to be differences in forecasting performance if the specific episodes had been isolated for comparisons over a shorter span of time.

Also of interest was the comparison of the NBER-ASA survey results with the econometric approach to forecasting as represented by the Michigan model (RSQE) forecasts. The Michigan model comparison with the NBER-ASA forecasts were unfavorable and did not support some earlier work that has indicated "model plus judgment" as superior to "judgmental" forecasting alone. The NBER-ASA survey showed about half or more of individual forecasts to be at least somewhat more accurate than Michigan in the major variables real GNP, inflation, and unemployment.

The comparison of the NBER-ASA surveys with the Michigan model really does not put the econometric approach in the best light, however. The Michigan model was chosen principally because it had the longest forecast history of those methods that were econometrics based, matching better the span of the NBER-ASA surveys. The more general system methods for macro forecasting, pioneered by Lawrence R. Klein and the various Klein and Wharton models and by Otto Eckstein (1983) at Data Resources, probably were not well represented by the Michigan model.

The National Economic Information System, Eckstein's terminology, is used more or less by many large-scale macroeconometric model builders, especially commercial vendors, including Data Resources, Wharton Econometric Forecasting Associates (WEFA), Evans Economics, Lawrence Meyer and Associates, and the Boston Company Economic Advisors. It stresses intensive monitoring, screening, and filtering of high-frequency data in the forecasts of early periods, typically spends more time assessing and estimating exogenous variables and policy inputs, and applies significant numbers of staff to the forecasting task. The National Economic Information System approach also encompasses many more types of studies, simulations, and collateral research that feed into the macroeconometric forecasts than perhaps might be followed by Michigan.

An interesting comparison would be to take macroeconometric model forecasts including those of the Michigan RSQE, perhaps the data kept by Steve McNees from the NBER-NSF Model Comparison Seminars, and to compare the collective results of those forecasts over a sample period that matches a portion of the NBER-ASA surveys. This would produce a more useful and informative comparison of the econometric approach with informal judgment (the NBER-ASA survey) than ZB actually performed. Group mean forecast comparisons between the two, NBER-ASA and econometrics based, would provide the most instructive test.

Presumably, these organizations who use the econometric approach in profit-making enterprises ought to produce, on average, more accurate forecasts than the NBER-ASA survey. The fees paid by the market are supposed to reflect some added value, perhaps related to the accuracy of the forecasting.

An important contribution of ZB was the notion of assessing accuracy relative to the variability of what is being predicted, an important point often neglected in the evaluation of forecasts and forecasting performance and in model building. Looking at forecast errors relative to actual changes or prospective changes, instead of actual levels, gives an idea of how much one can reasonably expect to forecast and a way to judge forecasts. Some variables are just inherently less easily forecast than others. In such a situation, forecasts that miss more could actually be better than forecasts of variables that follow an easily predicted path, or more accurate, or more valuable for decisions.

### Some Lessons for the Forecasting Practitioner

A few lessons or "messages" for forecasting practitioners emerge from the ZB work. First, forecast errors grow larger the longer the time span of a projection. All variables in the individual forecasts showed this property, as did also the group mean or consensus forecasts. This is neither a lesson nor a surprise, just a reaffirmation of how hard it is to make conditional forecasts of any variable with any method going out a long period of time because of events, right-hand-side variable perturbation, and noise between the time of the forecast and the time of the forecast realization. The lesson for the practicing forecaster is to forecast continuously or to develop systems that do so, monitoring external impulses and internal propagation mechanisms that might move variables of interest on a high-frequency basis.

Second, there is the finding, very well documented in ZB, that group mean forecasts (consensus) outperform most individual forecasts. Combining forecasts that come from different sources or that are generated by different techniques tends to produce significant gains in accuracy. This finding, at least in ZB, is accompanied by the notion that it is invariant to the forecast horizon and other time-series methods. Some individual forecasts do better than the consensus, about 25 percent of them for nominal GNP, real growth, and the implicit GNP deflator. For the practitioner who might use forecasts of inputs in his or her own forecast, this suggests averaging the sources of those inputs or generating several attempts at them by using different forecasting methods.

Third, there is much evidence relating to comparisons between forecasting methods. The methods included those used in the NBER-ASA survey itself, a potpourri of methods that most in the survey indicated as informal judgment, the BVAR approach as specified by ZB, a more sophisticated BVAR model provided by Sims, the Sims-Todd ARIMA, and the Michigan model (the "model-plus-manager" approach). The results (tables 1.26–1.31), measured by average root mean square error (ARMSE), tended to favor the informal GNP approach represented by the group mean forecasts of the NBER-ASA survey. Some time-series approaches performed impressively, particularly the Sims BVAR (variant B) and the ZB form of the BVAR (variant B). Other results were mixed, depending on the sample period and variables examined. The Michigan model generally did well in comparison with the various time-series methods.

A lesson here is the running of BVAR forecasting models as a collateral system, using some current-quarter information. The BVAR results are worth noting. Establishing such systems as potential generators of forecasts should be tried. The structural econometric model approach of the Michigan model showed favorably compared to most other methods except the group mean forecasts of the NBER-ASA. Along with the NBER-ASA and BVAR forecasts, the econometric method of forecasting ranked well. It is clear that macroeconomic forecasting cannot yet be expected to produce better results through any "black box" approach, regardless of the technique. The ARIMA results tended to rank lowest. BVAR forecasting, although clearly worthy of use at least as a collateral information input, consistently ranked below the group mean forecasts of the NBER-ASA and did no better than the econometric approach.

A fourth lesson relates to systematic under- or overprediction of macroeconomic time series. This is especially true of forecasts just after turning points, a topic not covered by ZB, with carryover in a time-series sense a prevalent characteristic. This tendency has actually been long known in forecasting. The implication for forecasters is to adjust forecasts from knowledge of this tendency, assuming that it will exist and persist. Many users of macro forecasts adjust consensus forecasts with this in mind, particularly the projections of financial market participants.

Fifth, judging accuracy by some standard of what is inherently characteristic of the series to be predicted is suggested by relating forecast changes to actual changes. If there is a lot of variability, then the standard for accuracy can be relaxed compared with cases where the variable being forecast has a smoother pattern. Validation in model building and forecast evaluation can make use of this notion.

## What Users of Forecasts Can Learn

There is much in ZB of use to consumers of macroeconomic forecasts, who, in decision making, can make adjustments to the forecasts that are provided in order to understand, interpret, and use them in a practical way. First, the cumulation of errors across the forecast horizon should make users skeptical of long-term forecasts, basically disbelieving them, and understanding this as the rule rather than the exception. Indeed, *forecast* is probably the wrong word to describe long-run projections that are conditional on so many events that can change between the time the forecast is made and the realization. The long-run projections should be taken as *scenarios* or planning paths.

Second, the use of consensus forecasts or averages from many sources can be a valuable informational input for planning. The superior accuracy of the group mean forecasts was impressive, widespread, and one of the significant themes in the ZB paper. Users of macroeconomic forecasts should put this notion to good use, averaging the forecasts of individual forecasters as one information input or averaging projections generated by different methods of forecasting. However, the group mean results of the NBER-ASA survey did show that about 25 percent of the individual forecasts were superior to the consensus, so it should be noted that significant value can be obtained from certain individual forecasts if accuracy is the main criterion.

Third, while time-series models are perhaps a useful adjunct or alternative informational input, they do not seem to offer a great deal in terms of accuracy. Users need to be skeptical of time-series or black box kinds of statistical projections. At the same time, an informationally dense BVAR approach seems worthy of note.

Finally, users should understand the systematic nature, either under- or overprediction, in macroeconomic forecasts, documented by ZB, and incorporate the systematic serially autocorrelated errors of macroeconomic forecasts into planning.

**Some Additional Observations**

There are some myths in the ZB paper. One has to do with the notion that forecasting macroeconomic aggregates like real GNP and its components has less potential for direct profitability than forecasting financial variables. I would not say this. In financial market work, there is room for the heavy use of forecasts of macro aggregates and their components, not as financial variables in themselves, but as inputs that will drive financial markets in a predictable way. The macro aggregates also can have implications for inflation, monetary policy, the trade balance, and currencies. As an example, the Federal Reserve drives markets through changes in short-term interest rates. Actual interest rate forecasts can be irrelevant. Understanding what may happen to the inputs that affect Fed policy can be much more valuable. Indeed, good forecasts of the macro aggregates can be more forward looking than the information contained in forecasts of interest rates, making the macro input very important. Virtually all decision makers use macro aggregates as a backdrop for planning, implicitly or explicitly.

A second mistaken notion relates to the assertion that the main value of the forecast lies in the ability to reduce uncertainty about the future faced by the user. Forecasts are a dime a dozen. Probably more valuable than a point forecast is a process that enhances understanding of the phenomenon being forecast or a clear statement of the way the forecasts are being generated.

## Conclusion

The ZB paper is an important and significant contribution to the literature on macroeconomic forecasting, with a valuable set of data, not even fully analyzed—and not without deficiencies—but probably the best set of data around.

The ZB record of the history of forecasting performance, insights into tendencies from the forecasts, comparison of the NBER-ASA informal model results with other forecasting methods, and striking observations that averages of individual forecasts are more accurate than individual forecasts alone constitute a major contribution of a paper that carefully uses a rich data set to provide a basic reference. The ZB contribution will be an essential reference document on macroeconomic forecasts and forecasting.

## References

Braun, P., S. Shishado, and J. Vuchelin, eds. Monthly. *Economic forecasts: A monthly worldwide survey.* Amsterdam: North-Holland.

Eckstein, Otto. 1983. *The DRI model of the U.S. economy.* New York: McGraw-Hill.

McNees, Stephen K. 1979. The forecasting record for the 1970s. *New England Economic Review* (September/October), 33–53.

———. 1981. The recent record of thirteen forecasters. *New England Economic Review* (September/October), 5–21.

———. 1988. How accurate are macroeconomic forecasts? *New England Economic Review* (July/August), 15–36.

McNees, Stephen K., and John Ries. 1983. The track record of macroeconomic forecasts. *New England Economic Review* (November/December), 5–18.

This Page Intentionally Left Blank